

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/85, 15/86, 7/01, A61K 48/00, C12N 5/10		A2	(11) International Publication Number: WO 95/14100
			(43) International Publication Date: 26 May 1995 (26.05.95)
(21) International Application Number: PCT/GB94/02546 (22) International Filing Date: 18 November 1994 (18.11.94) (30) Priority Data: 154,712 19 November 1993 (19.11.93) US (71) Applicant (for all designated States except US): THE WELL- COME FOUNDATION LIMITED [GB/GB]; Unicorn House, 160 Euston Road, London NW1 2BP (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): RICHARDS, Cynthia, Ann [US/US]; 5616 Welkin Court, Durham, NC 27713 (US). HUBER, Brian [US/US]; 53 Westridge Drive, Durham, NC 27713 (US). (74) Agent: STOTT, Michael, John; The Wellcome Foundation Limited, Langley Court, Beckenham, Kent BR3 3BS (GB).		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ). Published Without international search report and to be republished upon receipt of that report.	
(54) Title: TRANSCRIPTIONAL REGULATORY SEQUENCE OF CARCINOEMBRYONIC ANTIGEN FOR EXPRESSION TARGET- ING			
<p style="text-align: center;">λCEA1</p>			
(57) Abstract The invention relates to the transcriptional regulatory sequence (TRS) of carcinoembryonic antigen (CEA) and molecular chimera comprising the CEA TRS and DNA encoding a heterologous enzyme. CEA TRS is capable of targeting expression of the heterologous enzyme to CEA ⁺ cells and the heterologous enzyme is preferably an enzyme capable of catalysing the production of an agent cytotoxic or cytostatic to CEA ⁺ cells. For example the enzyme may be cytosine deaminase which is capable of catalysing formation of the cytotoxic compound 5-fluorouracil from the non toxic compound 5-fluorocytosine.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

**TRANSCRIPTIONAL REGULATORY SEQUENCE OF CARCINOEMBRYONIC ANTIGEN FOR
EXPRESSION TARGETING**

The present invention relates to a transcriptional regulatory sequence useful in gene therapy.

Colorectal carcinoma (CRC) is the second most frequent cancer and the second leading cause of cancer-associated deaths in the United States and Western Europe. The overall five-year survival rate for patients has not meaningfully improved in the last three decades. Prognosis for the CRC cancer patient is associated with the depth of tumor penetration into the bowel wall, the presence of regional lymph node involvement and, most importantly, the presence of distant metastases. The liver is the most common site for distant metastasis and, in approximately 30% of patients, the sole initial site of tumor recurrence after successful resection of the primary colon cancer. Hepatic metastases are the most common cause of death in the CRC cancer patient.

The treatment of choice for the majority of patients with hepatic CRC metastasis is systemic or regional chemotherapy using 5-fluorouracil (5-FU) alone or in combination with other agents such as levamisole. However, despite extensive effort, there is still no satisfactory treatment for hepatic CRC metastasis. Systemic single- and combination-agent chemotherapy and radiation are relatively ineffective emphasizing the need for new approaches and therapies for the treatment of the diseases.

A gene therapy approach is being developed for primary and metastatic liver tumors that exploits the transcriptional differences between normal and metastatic cells. This approach involves linking the transcriptional regulatory sequences (TRS) of a tumor associated marker gene to the coding sequence of an enzyme, typically a non-

mammalian enzyme, to create an artificial chimaeric gene that is selectively expressed in cancer cells. The enzyme should be capable of converting a non-toxic prodrug into a cytotoxic or cytostatic drug thereby allowing for selective
5 elimination of metastatic cells.

The principle of this approach has been demonstrated using an alpha-fetoprotein/Varicella Zoster virus thymidine kinase chimaera to target hepatocellular carcinoma with the
10 enzyme metabolically activating the non-toxic prodrug 6-methoxypurine arabinonucleoside ultimately leading to formation of the cytotoxic anabolite adenine arabinonucleoside triphosphate (see Huber *et al*, Proc. Natl. Acad. Sci U.S.A., 88, 8039-8043 (1991) and EP-A-0 415
15 731).

For the treatment of hepatic metastases of CRC, it is desirable to control the expression of an enzyme with the transcriptional regulatory sequences of a tumor marker
20 associated with such metastases.

CEA is a tumor associated marker that is regulated at the transcriptional level and is expressed by most CRC tumors but is not expressed in normal liver. CEA is widely
25 used as an important diagnostic tool for postoperative surveillance, chemotherapy efficacy determinations, immunolocalisation and immunotherapy. The TRS of CEA are potentially of value in the selective expression of an enzyme in CEA tumor cells since there appears to be a very
30 low heterogeneity of CEA within metastatic tumors, perhaps because CEA may have an important functional role in metastasis.

The cloning of the CEA gene has been reported and the
35 promoter localised to a region of 424 nucleotides upstream from the translational start (Schrewe *et al*, Mol. Cell. Biol., 10, 2738 - 2748 (1990) but the full TRS was not

identified.

In the work on which the present invention is based, CEA genomic clones have been identified and isolated from the human chromosome 19 genomic library LL19NL01, ATCC number 57766, by standard techniques described hereinafter. The cloned CEA sequences comprise CEA enhancers in addition to the CEA promoter. The CEA enhancers are especially advantageous for high level expression in CEA-positive cells and no expression in CEA-negative cells.

According to one aspect, the present invention provides a DNA molecule comprising the CEA TRS but without associated CEA coding sequence.

According to another aspect, the present invention provides use of a CEA TRS for and targeting expression of a heterologous enzyme to CEA⁺ cells. Preferably the enzyme is capable of catalysing the production of an agent cytotoxic or cytostatic to the CEA⁺ target cells.

As described in more detail hereinafter, the present inventors have sequenced a large part of the CEA gene upstream of the coding sequence. As used herein, the term "CEA TRS" means any part of the CEA gene upstream of the coding sequence which has a transcriptional regulatory effect on a heterologous coding sequence operably linked thereto.

Certain parts of the sequence of the CEA gene upstream of the coding sequence have been identified as making significant contributions to the transcriptional regulatory

3(2/2)

effect, more particularly increasing the level and/or selectivity of transcription.

Preferably the CEA TRS includes all or part of the region from about -299b to about +69b, more preferably about -90b to about +69b. Increases in the level of transcription and/or selectivity can also be obtained by including one or more of the following regions: -14.5kb to -10.6kb, preferably -13.6kb to -10.6kb, and/or -6.1kb to -3.8kb. All of the regions referred to above can be included in either orientation and in different combinations. In addition, repeats of these regions may be included, particularly repeats of the -90b to +69b region, containing for example 2,3, 4 or more copies of the region. The base numbering refers to the sequence of Figure 6. The regions referred to are included in the plasmids described in figure 5B.

Gene therapy involves the stable integration of new genes into target cells and the expression of those genes, once they are in place, to alter the phenotype of that particular target cell (for review see Anderson, W.F. *Science* 226, 401-409 (1984) and McCormick, D. *Biotechnology* 3, 689-693, (1985)). Gene therapy may be beneficial for the treatment of genetic diseases that involve the replacement of one defective or missing enzyme, such as; hypoxanthine-guanine phosphoribosyl transferase in Lesch-Nyhan disease, purine nucleoside phosphorylase in severe immunodeficiency disease, and adenosine deaminase in severed combined immunodeficiency disease.

It has now been found that it is possible to selectively arrest the growth of, or kill, mammalian carcinoma cells with prodrugs, i.e. chemical agents capable

of selective conversion to cytotoxic (causing cell death) or cytostatic (suppressing cell multiplication and growth) metabolites. This is achieved by the construction of a molecular chimaera comprising a "target tissue-specific" TRS that is selectively activated in target cells, such as cancerous cells, and that controls the expression of a heterologous enzyme. This molecular chimaera may be manipulated via suitable vectors and incorporated into an infective virion. Upon administration of an infective virion containing the molecular chimaera to a host (e.g., mammal or human), the enzyme is selectively expressed in the target cells. Administration of prodrugs (compounds that are selectively metabolised by the enzyme into metabolites that are either further metabolised to or are, in fact, cytotoxic or cytostatic agents) can then result in the production of the cytotoxic or cytostatic agent in situ in the cancer cell. According to the present invention CEA TRS provides the target tissue specificity.

Molecular chimaeras (recombinant molecules comprised of unnatural combinations of genes or sections of genes), and infective virions (complete viral particles capable of infecting appropriate host cells) are well known in the art of molecular biology.

A number of enzyme prodrug combinations may be used for the above purpose, providing the enzyme is capable of selectively activating the administered compound either directly or through an intermediate to a cytostatic or cytotoxic metabolite. The choice of compound will also depend on the enzyme system used, but must be selectively metabolised by the enzyme either directly or indirectly to a cytotoxic or cytostatic metabolite. The term heterologous enzyme, as used herein, refers to an enzyme that is derived from or associated with a species which is different from the host to be treated and which will display the appropriate characteristics of the above

mentioned selectivity. In addition, it will also be appreciated that a heterologous enzyme may also refer to an enzyme that is derived from the host to be treated that has been modified to have unique characteristics unnatural to the host.

The enzyme cytosine deaminase (CD) catalyses the deamination of cytosine to uracil. Cytosine deaminase is present in microbes and fungi but absent in higher eukaryotes. This enzyme catalyses the hydrolytic deamination of cytosine and 5-fluorocytosine (5-FC) to uracil and 5-fluorouracil (5-FU), respectively. Since mammalian cells do not express significant amounts of cytosine deaminase, they are incapable of converting 5-FC to the toxic metabolite 5-FU and therefore 5-fluorocytosine is nontoxic to mammalian cells at concentrations which are effective for antimicrobial activity. 5-Fluorouracil is highly toxic to mammalian cells and is widely used as an anticancer agent.

In mammalian cells, some genes are ubiquitously expressed. Most genes, however, are expressed in a temporal and/or tissue-specific manner, or are activated in response to extracellular inducers. For example, certain genes are actively transcribed only at very precise times in ontogeny in specific cell types, or in response to some inducing stimulus. This regulation is mediated in part by the interaction between transcriptional regulatory sequences (for example, promoter and enhancer regulatory DNA sequences), and sequence-specific, DNA-binding transcriptional protein factors.

It has now been found that it is possible to alter certain mammalian cells, e.g. colorectal carcinoma cells, metastatic colorectal carcinoma cells and hepatic colorectal carcinoma cells to selectively express a heterologous enzyme as hereinbefore defined, e.g. CD. This

is achieved by the construction of molecular chimaeras in an expression cassette.

5 Expression cassettes themselves are well known in the art of molecular biology. Such an expression cassette contains all essential DNA sequences required for expression of the heterologous enzyme in a mammalian cell. For example, a preferred expression cassette will contain a molecular chimaera containing the coding sequence for CD,
10 an appropriate polyadenylation signal for a mammalian gene (i.e., a polyadenylation signal that will function in a mammalian cell), and CEA enhancers and promoter sequences in the correct orientation.

15 Normally, two DNA sequences are required for the complete and efficient transcriptional regulation of genes that encode messenger RNAs in mammalian cells: promoters and enhancers. Promoters are located immediately upstream (5') from the start site of transcription. Promoter
20 sequences are required for accurate and efficient initiation of transcription. Different gene-specific promoters reveal a common pattern of organisation. A typical promoter includes an AT-rich region called a TATA box (which is located approximately 30 base pairs 5' to the
25 transcription initiation start site) and one or more upstream promoter elements (UPEs). The UPEs are a principle target for the interaction with sequence-specific nuclear transcriptional factors. The activity of promoter sequences is modulated by other sequences called enhancers.
30 The enhancer sequence may be a great distance from the promoter in either an upstream (5') or downstream (3') position. Hence, enhancers operate in an orientation- and position-independent manner. However, based on similar structural organisation and function that may be
35 interchanged, the absolute distinction between promoters and enhancers is somewhat arbitrary. Enhancers increase the rate of transcription from the promoter sequence. It

is predominantly the interaction between sequence-specific transcriptional factors with the UPE and enhancer sequences that enable mammalian cells to achieve tissue-specific gene expression. The presence of these transcriptional protein factors (tissue-specific, trans-activating factors) bound to the UPE and enhancers (cis-acting, regulatory sequences) enables other components of the transcriptional machinery, including RNA polymerase, to initiate transcription with tissue-specific selectivity and accuracy.

10

The transcriptional regulatory sequence for CEA is suitable for targeting expression in colorectal carcinoma, metastatic colorectal carcinoma, and hepatic colorectal metastases, transformed cells of the gastrointestinal tract, lung, breast and other tissues. By placing the expression of the gene encoding CD under the transcriptional control of the CRC-associated marker gene, CEA, the nontoxic compound, 5-FC, can be metabolically activated to 5-FU selectively in CRC cells (for example, hepatic CRC cells). An advantage of this system is that the generated toxic compound, 5-fluorouracil, can diffuse out of the cell in which it was generated and kill adjacent tumor cells which did not incorporate the artificial gene for cytosine deaminase.

25

In the work on which the present invention is based, CEA genomic clones were identified and isolated from the human chromosome 19 genomic library LL19NL01, ATCC number 57766, by standard techniques described hereinafter. The cloned CEA sequences comprise CEA enhancers in addition to the CEA promoter. The CEA enhancers are especially advantageous for high level expression in CEA-positive cells and no expression in CEA-negative cells.

35

The present invention further provides a molecular chimaera comprising a CEA TRS and a DNA sequence operatively linked thereto encoding a heterologous enzyme,

preferably an enzyme capable of catalysing the production of an agent cytotoxic or cytostatic to the CEA⁺ cells.

5 The present invention further provides a molecular chimaera comprising a DNA sequence containing the coding sequence of the gene that codes for a heterologous enzyme under the control of a CEA TRS in an expression cassette.

10 The present invention further provides in a preferred embodiment a molecular chimaera comprising a CEA TRS which is operatively linked to the coding sequence for the gene encoding a non-mammalian cytosine deaminase (CD). The molecular chimaera comprises a promoter and additionally comprises an enhancer.

15 In particular, the present invention provides a molecular chimaera comprising a DNA sequence of the coding sequence of the gene coding for the heterologous enzyme, which is preferably CD, additionally including an appropriate polyadenylation sequence, which is linked downstream in a 3' position and in the proper orientation to a CEA TRS. Most preferably the expression cassette also contains an enhancer sequence.

20 Preferably non-mammalian CD is selected from the group consisting of bacterial, fungal, and yeast cytosine deaminase.

25 The molecular chimaera of the present invention may be made utilizing standard recombinant DNA techniques.

30 Another aspect of the invention is the genomic CEA sequence as described by Seq ID1.

35 The coding sequence of CD and a polyadenylation signal (for example see Seq IDs 1 and 2) are placed in the proper 3' orientation to the essential CEA transcriptional

regulatory elements. This molecular chimaera enables the selective expression of CD in cells or tissue that normally express CEA. Expression of the CD gene in mammalian CRC and metastatic CRC (hepatic colorectal carcinoma metastases) will enable nontoxic 5-FC to be selectively metabolised to cytotoxic 5-FU.

Accordingly, in a another aspect of the present invention, there is provided a method of constructing a molecular chimaera comprising linking a DNA sequence encoding a heterologous enzyme gene, e.g. CD, to a CEA TRS.

In particular the present invention provides a method of constructing a molecular chimaera as herein defined, the method comprising ligating a DNA sequence encoding the coding sequence and polyadenylation signal of the gene for a heterologous enzyme (e.g. non-mammalian CD) to a CEA TRS (e.g., promoter sequence and enhancer sequence).

These molecular chimaeras can be delivered to the target tissue or cells by a delivery system. For administration to a host (e.g., mammal or human), it is necessary to provide an efficient in vivo delivery system that stably incorporates the molecular chimaera into the cells. Known methods utilize techniques of calcium phosphate transfection, electroporation, microinjection, liposomal transfer, ballistic barrage, DNA viral infection or retroviral infection. For a review of this subject see Biotechniques 6, No.7, (1988).

The technique of retroviral infection of cells to integrate artificial genes employs retroviral shuttle vectors which are known in the art (Miller A.D., Buttimore C. Mol. Cell. Biol. 6, 2895-2902 (1986)). Essentially, retroviral shuttle vectors (retroviruses comprising molecular chimaeras used to deliver and stably integrate the molecular chimaera into the genome of the target cell)

are generated using the DNA form of the retrovirus contained in a plasmid. These plasmids also contain sequences necessary for selection and growth in bacteria. Retroviral shuttle vectors are constructed using standard
5 molecular biology techniques well known in the art. Retroviral shuttle vectors have the parental endogenous retroviral genes (e.g., gag, pol and env) removed from the vectors and the DNA sequence of interest is inserted, such as the molecular chimaeras that have been described. The
10 vectors also contain appropriate retroviral regulatory sequences for viral encapsidation, proviral insertion into the target genome, message splicing, termination and polyadenylation. Retroviral shuttle vectors have been derived from the Moloney murine leukaemia virus (Mo-MLV)
15 but it will be appreciated that other retroviruses can be used such as the closely related Moloney murine sarcoma virus. Other DNA viruses may also prove to be useful as delivery systems. The bovine papilloma virus (BPV) replicates extrachromosomally, so that delivery systems
20 based on BPV have the advantage that the delivered gene is maintained in a nonintegrated manner.

Thus according to a further aspect of the present invention there is provided a retroviral shuttle vector
25 comprising the molecular chimaeras as hereinbefore defined.

The advantages of a retroviral-mediated gene transfer system are the high efficiency of the gene delivery to the
30 targeted tissue or cells, sequence specific integration regarding the viral genome (at the 5' and 3' long terminal repeat (LTR) sequences) and little rearrangements of delivered DNA compared to other DNA delivery systems.

35 Accordingly in a preferred embodiment of the present invention there is provided a retroviral shuttle vector comprising a DNA sequence comprising a 5' viral LTR

sequence, a cis-acting psi-encapsidation sequence, a molecular chimaera as hereinbefore defined and a 3' viral LTR sequence.

5 In a preferred embodiment, and to help eliminate non-tissue-specific expression of the molecular chimaera, the molecular chimaera is placed in opposite transcriptional orientation to the 5' retroviral LTR. In addition, a dominant selectable marker gene may also be included that
10 is transcriptionally driven from the 5' LTR sequence. Such a dominant selectable marker gene may be the bacterial neomycin-resistance gene NEO (aminoglycoside 3' phospho-transferase type II), which confers on eukaryotic cells
15 resistance to the neomycin analogue Geneticin (antibiotic G418 sulphate; registered trademark of GIBCO). The NEO gene aids in the selection of packaging cells that contain these sequences.

 The retroviral vector is preferably based on the
20 Moloney murine leukaemia virus but it will be appreciated that other vectors may be used. Vectors containing a NEO gene as a selectable marker have been described, for example, the N2 vector (Eglitis M.A., Kantoff P., Gilboa E., Anderson W.F. Science 230, 1395-1398 (1985)).

25 A theoretical problem associated with retroviral shuttle vectors is the potential of retroviral long terminal repeat (LTR) regulatory sequences transcriptionally activating a cellular oncogene at the
30 site of integration in the host genome. This problem may be diminished by creating SIN vectors. SIN vectors are self-inactivating vectors that contain a deletion comprising the promoter and enhancer regions in the retroviral LTR. The LTR sequences of SIN vectors do not
35 transcriptionally activate 5' or 3' genomic sequences. The transcriptional inactivation of the viral LTR sequences diminishes insertional activation of adjacent target cell

DNA sequences and also aids in the selected expression of the delivered molecular chimaera. SIN vectors are created by removal of approximately 299 bp in the 3' viral LTR sequence (Gilboa E., Egilitis P.A., Kantoff P.W., Anderson W.F. Biotecniques 4, 504-512 (1986)).

Thus preferably the retroviral shuttle vectors of the present invention are SIN vectors.

10 Since the parental retroviral cag, pol, and env genes have been removed from these shuttle vectors, a helper virus system may be utilised to provide the cag, pol, and env retroviral gene products in trans to package or encapsidate the retroviral vector into an infective virion.

15 This is accomplished by utilising specialised "packaging" cell lines, which are capable of generating infectious, synthetic virus yet are deficient in the ability to produce any detectable wild-type virus. In this way the artificial synthetic virus contains a chimaera of the present

20 invention packaged into synthetic artificial infectious virions free of wild-type helper virus. This is based on the fact that the helper virus that is stably integrated into the packaging cell contains the viral structural genes, but is lacking the psi-site, a cis-acting regulatory

25 sequence which must be contained in the viral genomic RNA molecule for it to be encapsidated into an infectious viral particle.

Accordingly, in a still further aspect of the present

30 invention, there is provided an infective virion comprising a retroviral shuttle vector, as hereinbefore described, said vector being encapsidated within viral proteins to create an artificial, infective, replication-defective, retrovirus.

35

In a another aspect of the present invention there is provided a method for producing infective virions of the

present invention by delivering the artificial retroviral shuttle vector comprising a molecular chimera of the invention, as hereinbefore described, into a packaging cell line.

5

The packaging cell line may have stably integrated within it a helper virus lacking a psi-site and other regulatory sequence, as hereinbefore described, or, alternatively, the packaging cell line may be engineered so as to contain helper virus structural genes within its genome. In addition to removal of the psi-site, additional alterations can be made to the helper virus LTR regulatory sequences to ensure that the helper virus is not packaged in virions and is blocked at the level of reverse transcription and viral integration. Alternatively, helper virus structural genes (i.e., gag, pol, and env) may be individually and independently transferred into the packaging cell line. Since these viral structural genes are separated within the packaging cell's genome, there is little chance of covert recombinations generating wild-type virus.

10
15
20

The present invention also provides a packaging cell line comprising an infective virion, as described hereinbefore, said virion further comprising a retroviral shuttle vector.

25

The present invention further provides for a packaging cell line comprising a retroviral shuttle vector as described hereinbefore.

30

In addition to retroviral-mediated gene delivery of the chimeric, artificial, therapeutic gene, other gene delivery systems known to those skilled in the art can be used in accordance with the present invention. These other gene delivery systems include other viral gene delivery systems known in the art, such as the adenovirus delivery

35

systems.

Non-viral delivery systems can be utilized in accordance with the present invention as well. For example, liposomal delivery systems can deliver the therapeutic gene to the tumor site via a liposome. Liposomes can be modified to evade metabolism and/or to have distinct targeting mechanisms associated with them. For example, liposomes which have antibodies incorporated into their structure, such as antibodies to CEA, can have targeting ability to CEA-positive cells. This will increase both the selectivity of the present invention as well as its ability to treat disseminated disease (metastasis).

Another gene delivery system which can be utilized according to the present invention is receptor-mediated delivery, wherein the gene of choice is incorporated into a ligand which recognizes a specific cell receptor. This system can also deliver the gene to a specific cell type. Additional modifications can be made to this receptor-mediated delivery system, such as incorporation of adenovirus components to the gene so that the gene is not degraded by the cellular lysosomal compartment after internalization by the receptor.

The infective virion or the packaging cell line according to the invention may be formulated by techniques well known in the art and may be presented as a formulation (composition) with a pharmaceutically acceptable carrier therefor. Pharmaceutically acceptable carriers, in this instance physiologic aqueous solutions, may comprise liquid medium suitable for use as vehicles to introduce the infective virion into a host. An example of such a carrier is saline. The infective virion or packaging cell line may be a solution or suspension in such a vehicle. Stabilizers and antioxidants and/or other excipients may also be

present in such pharmaceutical formulations (compositions), which may be administered to a mammal by any conventional method (e.g., oral or parenteral routes). In particular, the infective virion may be administered by intra-venous or intra-arterial infusion. In the case of treating hepatic metastatic CRC, intra-hepatic arterial infusion may be advantageous. The packaging cell line can be administered directly to the tumor or near the tumor and thereby produce infective virions directly at or near the tumor site.

10

Accordingly, the present invention provides a pharmaceutical formulation (composition) comprising an infective virion or packaging cell line according to the invention in admixture with a pharmaceutically acceptable carrier.

15

Additionally, the present invention provides methods of making pharmaceutical formulations (compositions), as herein described, comprising mixing an artificial infective virion, containing a molecular chimaera according to the invention as described hereinbefore, with a pharmaceutically acceptable carrier.

20

The present invention also provides methods of making pharmaceutical formulations (compositions), as herein described, comprising mixing a packaging cell line, containing an infective virion according to the invention as described hereinbefore, with a pharmaceutically acceptable carrier.

25

Although any suitable compound that can be selectively converted to a cytotoxic or cytostatic metabolite by the enzyme cytosine deaminase may be utilised, the preferred compound for use according to the invention is 5-FC, in particular for use in treating colorectal carcinoma (CRC), metastatic colorectal carcinoma, or hepatic CRC metastases. 5-FC, which is non-toxic and is used as an antifungal, is

30

35

converted by CD into the established cancer therapeutic 5-FU.

Any agent that can potentiate the antitumor effects of 5-FU can also potentiate the antitumor effects of 5-FC since, when used according to the present invention, 5-FC is selectively converted to 5-FU. According to another aspect of the present invention, agents such as leucovorin and levamisol, which can potentiate the antitumor effects of 5-FU, can also be used in combination with 5-FC when 5-FC is used according to the present invention. Other agents which can potentiate the antitumor effects of 5-FU are agents which block the metabolism 5-FU. Examples of such agents are 5-substituted uracil derivatives, for example, 5-ethynyluracil and 5-bromovinyluracil (PCT/GB91/01650 (WO 92/04901); Cancer Research 46, 1094, (1986) which are incorporated herein by reference in their entirety). Therefore, a further aspect of the present invention is the use of an agent which can potentiate the antitumor effects of 5-FU, for example, a 5-substituted uracil derivative such as 5-ethynyluracil or 5-bromovinyluracil in combination with 5-FC when 5-FC is used according to the present invention. The present invention further includes the use of agents which are metabolised in vivo to the corresponding 5-substituted uracil derivatives described hereinbefore (see Biochemical Pharmacology 38, 2885, (1989) which is incorporated herein by reference in its entirety) in combination with 5-FC when 5-FC is used according to the present invention.

30

5-FC is readily available (e.g., United States Biochemical, Sigma) and well known in the art. Leucovorin and levamisol are also readily available and well known in the art.

35

Two significant advantages of the enzyme/prodrug combination of cytosine deaminase/5-fluorocytosine and

further aspects of the invention are the following:

1. The metabolic conversion of 5-FC by CD produces 5-FU which is the drug of choice in the treatment of many different types of cancers, such as colorectal carcinoma.
2. The 5-FU that is selectively produced in one cancer cell can diffuse out of that cell and be taken up by both non-facilitated diffusion and facilitated diffusion into adjacent cells. This produces a neighbouring cell killing effect. This neighbour cell killing effect alleviates the necessity for delivery of the therapeutic molecular chimera to every tumor cell. Rather, delivery of the molecular chimera to a certain percentage of tumor cells can produce the complete eradication of all tumor cells.

The amounts and precise regimen in treating a mammal, will of course be the responsibility of the attendant physician, and will depend on a number of factors including the type and severity of the condition to be treated. However, for hepatic metastatic CRC, an intrahepatic arterial infusion of the artificial infective virion at a titer of between 2×10^5 and 2×10^7 colony forming units per ml (CFU/ml) infective virions is suitable for a typical tumour. Total amount of virions infused will be dependent on tumour size and are preferably given in divided doses.

Likewise, the packaging cell line is administered directly to a tumor in an amount of between 2×10^5 and 2×10^7 cells. Total amount of packaging cell line infused will be dependent on tumour size and is preferably given in divided doses.

Prodrug treatment - Subsequent to infection with the infective virion, certain cytosine compounds (prodrugs of 5-FU) are converted by CD to cytotoxic or cytostatic metabolites (e.g. 5-FC is converted to 5-FU) in target

cells. The above mentioned prodrug compounds are administered to the host (e.g. mammal or human) between six hours and ten days, preferably between one and five days, after administration of the infective virion.

5

The dose of 5-FC to be given will advantageously be in the range 10 to 500 mg per kg body weight of recipient per day, preferably 50 to 500 mg per kg bodyweight of recipient per day, more preferably 50 to 250 mg per kg bodyweight of recipient per day, and most preferably 50 to 150 mg per kg body weight of recipient per day. The mode of administration of 5-FC in humans are well known to those skilled in the art. Oral administration and/or constant intravenous infusion of 5-FC are anticipated by the instant invention to be preferable.

10
15

The doses and mode of administration of leucovorin and levamisole to be used in accordance with the present invention are well known or readily determined by those clinicians skilled in the art of oncology.

20

The dose and mode of administration of the 5-substituted uracil derivatives can be determined by the skilled oncologist. Preferably, these derivatives are given by intravenous injection or orally at a dose of between 0.01 to 50 mg per kg body weight of the recipient per day, particularly 0.01 to 10 mg per kg body weight per day, and more preferably 0.01 to 0.4 mg per kg bodyweight per day depending on the derivative used. An alternative preferred administration regime is 0.5 to 10 mg per kg body weight of recipient once per week.

25
30

The following examples serve to illustrate the present invention but should not be construed as a limitation thereof. In the Examples reference is made to the Figures a brief description of which is as follows:

35

Figure 1: Diagram of CEA phage clones. The overlapping clones lambdaCEA1, lambdaCEA7, and lambdaCEA5 represent an approximately 26 kb region of CEA genomic sequence. The 11,288 bp HindIII-Sau3A fragment that was sequenced is represented by the heavy line under lambdaCEA1. The 3774 bp HindIII-HindIII fragment that was sequenced is represented by the heavy line under lambdaCEA7. The bent arrows represent the transcription start point for CEA mRNA. The straight arrows represent the oligonucleotides CR15 and CR16. H, HindIII; S, SstI; B, BamHI; E, EcoRI; X, XbaI.

Figure 2: Restriction map of part of lambdaCEA1. The arrow head represents the approximate location of the transcription initiation point for CEA mRNA. Lines below the map represent the CEA inserts of pBS+ subclones. These subclones are convenient sources for numerous CEA restriction fragments.

DNA sequence of the 11,288 bp HindIII to Sau3A fragment of lambdaCEA7 (SEQ ID NO: 1). Sequence is numbered with the approximate transcription initiation point for CEA mRNA as 0 (this start site is approximate because there is some slight variability in the start site among individual CEA transcripts). The translation of the first exon is shown. Intron 1 extends from +172 to beyond +592. Several restriction sites are shown above the sequence. In subclone 109-3 the sequence at positions +70 has been altered by site-directed mutagenesis in order to introduce HindIII and EcoRI restriction sites.

DNA sequence of the 3774 bp Hind III to HindIII fragment of lambda CEA7 (SEQ ID NO: 2).

35

Figure 3 : Mapplot of 15,056 bp HindIII to Sau3A fragment from CEA genomic DNA showing consensus sequences.

Schematic representation of some of the consensus sequences found in the CEA sequence of Seq IDs 1 and 2. The consensus sequences shown here are from the transcriptional dictionary of Locker and Buzard (DNA Sequence 1, 3-11 (1990)). The lysozymal silencer is coded B13. The last line represents 90% homology to the topoisomerase II cleavage consensus.

Figure 4: Cloning scheme for CEA constructs extending from -299 bp to +69 bp.

Figure 5A: Cloning scheme for CEA constructs extending from -10.7 kb to +69 bp.

Figure 5B: Coordinates for CEA sequence present in several CEA/luciferase clones. CEA sequences were cloned into the multiple cloning region of pGL2-Basic (Promega Corp.) by standard techniques.

Figures 5C and 5D: Transient luciferase assays. Transient transfections and luciferase assays were performed in quadruplicate by standard techniques using DOTAP (Boehringer Mannheim, Indianapolis, IN, USA), luciferase assay system (Promega, Madison, WI, USA), and Dynatech luminometer (Chantilly, VA, USA). CEA-positive cell lines included LoVo (ATCC #CCL 229) and SW1463 (ATCC #CCL 234). CEA-negative cell lines included HuH7 and Hep3B (ATCC #HB 8064). C. Luciferase activity expressed as the percent of pGL2-Control plasmid activity. D. Luciferase activities of LoVo and SW1463 expressed as fold increase over activity in Hep3B.

Example 1

Construction of transcriptional regulatory sequence of carcinoembryonic antigen/cytosine deaminase molecular

chimaeraA) Cloning and isolation of the transcriptional regulatory
sequence of the carcinoembryonic antigen gene

5

CEA genomic clones were identified and isolated from the human chromosome 19 genomic library LL19NL01, ATCC #57766, by standard techniques (Richards et al., Cancer Research, 50, 1521-1527 (1990) which is herein incorporated by reference in its entirety). The CEA clones were identified by plaque hybridization to ³²P end-labelled oligonucleotides CR15 and CR16. CR15, 5'-CCCTGTGATCTCCAGGACAGCTCAGTCTC-3' (SEQ ID NO: 3), and CR16, 5'-GTTTCCTGAGTGATGTCTGTGTGCAATG-3' (SEQ ID NO: 4), hybridize to a 5' non-transcribed region of CEA that has little homology to other members of the CEA gene family. Phage DNA was isolated from three clones that hybridized to both oligonucleotide probes. Polymerase chain reaction, restriction mapping, and DNA sequence analysis confirmed that the three clones contained CEA genomic sequences. The three clones are designated lambdaCEA1, lambdaCEA5, and lambdaCEA7 and have inserts of approximately 13.5, 16.2, and 16.7 kb respectively. A partial restriction map of the three overlapping clones is shown in Figure 1.

25

Clone lambdaCEA1 was initially chosen for extensive analysis. Fragments isolated from lambdaCEA1 were subcloned using standard techniques into the plasmid pBS+ (Stratagene Cloning Systems, La Jolla, CA, USA) to facilitate sequencing, site-directed mutagenesis, and construction of chimeric genes. The inserts of some clones are represented in Figure 2. The complete DNA sequence of a 11,288 bp HindIII/Sau3A restriction fragment from lambdaCEA1 (

30

SEQ ID NO: 1) was determined by the dideoxy sequencing method using the dsDNA Cycle Sequencing System from Life Technologies, Inc. and multiple oligonucleotide primers. This sequence extends from -10.7 kb to +0.6 kb relative to

35

the start site of CEA mRNA. The sequence of 3774 base pair HindIII restriction fragment from lambdaCEA1 was also determined (SEQ ID NO: 2). This sequence extends from -14.5 kb to -10.7 kb relative to the start site of CEA mRNA. This HindIII fragment is present in plasmid pCR145.

To determine important transcriptional regulatory sequences various fragments of CEA genomic DNA are linked to a reporter gene such as luciferase or chloramphenicol acetyltransferase. Various fragments of CEA genomic DNA are tested to determine the optimized, cell-type specific TRS that results in high level reporter gene expression in CEA-positive cells but not in CEA-negative cells. The various reporter constructs, along with appropriate controls, are transfected into tissue culture cell lines that express high, low, or no CEA. The reporter gene analysis identifies both positive and negative transcriptional regulatory sequences. The optimized CEA-specific TRS is identified through the reporter gene analysis and is used to specifically direct the expression of any desired linked coding sequence, such as CD or VZV TK, in cancerous cells that express CEA. The optimized CEA-specific TRS, as used herein, refers to any DNA construct that directs suitably high levels of expression in CEA positive cells and low or no expression in CEA-negative cells. The optimized CEA-specific TRS consists of one or several different fragments of CEA genomic sequence or multimers of selected sequences that are linked together by standard recombinant DNA techniques. It will be appreciated by those skilled in the art that the optimized CEA-specific TRS may also include some sequences that are not derived from the CEA genomic sequences shown in Seq IDs 1 and 2. These other sequences may include sequences from adjoining regions of the CEA locus, such as sequences from the introns, or sequences further upstr am or downstream from the sequenced DNA shown in Seq IDs 1 and 2, or they could include transcriptional control elements from other genes that when linked to

selected CEA sequences result in the desired CEA-specific regulation.

5 The CEA sequence of Seq IDs 1 and 2 were computer
analyzed for characterized consensus sequences which have
been associated with gene regulation. Currently not enough
is known about transcriptional regulatory sequences to
accurately predict by sequence alone whether a sequence
will be functional. However, computer searches for
10 characterized consensus sequences can help identify
transcriptional regulatory sequences in uncharacterized
sequences since many enhancers and promoters consist of
unique combinations and spatial alignments of several
characterized consensus sequences as well as other
15 sequences. Since not all transcriptional regulatory
sequences have been identified and not all sequences that
are identical to characterized consensus sequences are
functional, such a computer analysis can only suggest
possible regions of DNA that may be functionally important
20 for gene regulation.

Some examples of the consensus sequences that are
present in the CEA sequence are shown in
Figure 3. Four copies of a lysozymal silencer consensus
25 sequences have been found in the CEA sequence. Inclusion of
one or more copies of this consensus sequence in the
molecular chimera can help optimize CEA-specific
expression. A cluster of topoisomerase II cleavage
consensus identified approximately 4-5 kb upstream of the
30 CEA transcriptional start suggest that this region of CEA
sequence may contain important transcriptional regulatory
signals that may help optimize CEA-specific expression.

35 The first fragment of CEA genomic sequence analyzed for
transcriptional activity extends from -299 to +69, but it
is appreciated by those skilled in the art that other
fragments are tested in order to isolate a TRS that directs

strong expression in CEA-positive cells but little expression in CEA-negative cells. As diagrammed in Figure 4 the 943 bp SmaI-HindIII fragment of plasmid 39-5-5 was subcloned into the SmaI-HindIII sites of vector pBS+ (Statagene Cloning Systems) creating plasmid 96-11. Single-stranded DNA was rescued from cultures of XL1-blue 96-11 using an M13 helper virus by standard techniques. O l i g o n u c l e o t i d e C R 7 0 , 5 ' - CCTGGAAGCTCAAGCTTGAATTCTCCACAGAGGAGG-3' (SEQ ID NO: 5), was used as a primer for oligonucleotide-directed mutagenesis to introduce HindIII and EcoRI restriction sites at +65. Clone 109-3 was isolated from the mutagenesis reaction and was verified by restriction and DNA sequence analysis to contain the desired changes in the DNA sequence. CEA genomic sequences -299 to +69, original numbering Figure 3, were isolated from 109-3 as a 381 bp EcoRI/HindIII fragment. Plasmid pRc/CMV (Invitrogen Corporation, San Diego, CA, USA) was restricted with AatII and HindIII and the 4.5 kb fragment was isolated from low melting point agarose by standard techniques. The 4.5 kb fragment of pRc/CMV was ligated to the 381 bp fragment of 109-3 using T4 DNA ligase. During this ligation the compatible HindIII ends of the two different restriction fragments were ligated. Subsequently the ligation reaction was supplemented with the four deoxynucleotides, dATP, dCTP, dGTP, and dTTP, and T4 DNA polymerase in order to blunt the non-compatible AatII and EcoRI ends. After incubating, phenol extracting, and ethanol precipitating the reaction, the DNAs were again incubated with T4 DNA ligase. The resulting plasmid, pCR92, allows the insertion of any desired coding sequence into the unique HindIII site downstream of the CEA TRS, upstream from a polyadenylation site and linked to a dominant selectable marker. The coding sequence for CD or other desirable effector or reporter gene, when inserted in the correct orientation into the HindIII site, are transcriptionally regulated by the CEA sequences and are preferably expressed in cells

that express CEA but not in cells that do not express CEA.

In order to determine the optimized CEA TRS other reporter gene constructs containing various fragments of CEA genomic sequences are made by standard techniques from DNA isolated from any of the CEA genomic clones (Figures 1, 2, 4, and 5). DNA fragments extending from the HindIII site introduced at position +65 (original numbering Figure 3A) and numerous different upstream sites are isolated and cloned into the unique HindIII site in plasmid pSVOALdelta5' (De Wet, J.R., *et al.* Mol. Cell. Biol., 7, 725-737 (1987) which is herein incorporated by reference in its entirety) or any similar reporter gene plasmid to construct luciferase reporter gene constructs, Figures 4 and 5. These and similar constructs are used in transient expression assays performed in several CEA-positive and CEA-negative cell lines to determine a strong, CEA-positive cell-type specific TRS. Figures 5B, 5C, and 5D show the results obtained from several CEA/luciferase reporter constructs. The optimized TRS is used to regulate the expression of CD or other desirable gene in a cell-type specific pattern in order to be able to specifically kill cancer cells. The desirable expression cassette is added to a retroviral shuttle vector to aid in delivery of the expression cassette to cancerous tissue.

Strains containing plasmids 39-5-5 and 39-5-2 were deposited at the ATCC under the Budapest Treaty with Accession No. 68904 and 68905, respectively. A strain containing plasmid pCR92 was deposited with the ATCC under the Budapest Treaty with Accession No. 68914. A strain containing plasmid pCR145 was deposited at the ATCC under the Budapest Treaty with Accession No. 69460.

B) Cloning and isolation of the E. coli gene encoding cytosine deaminase (CD)

The cloning, sequencing and expression of *E. coli* CD has already been published (Austin & Huber, Molecular Pharmacology, 42, 380 - 387 (1993) the disclosure of which is incorporated herein by reference). A positive genetic selection was designed for the cloning of the *codA* gene from *E. coli*. The selection took advantage of the fact that *E. coli* is only able to metabolize cytosine via CD. Based on this, an *E. coli* strain was constructed that could only utilize cytosine as a pyrimidine source when cytosine deaminase was provided in *trans*. This strain, BA101, contains a deletion of the *codAB* operon and a mutation in the *pyrF* gene. The strain was created by transducing a *pyrF* mutation (obtained from the *E. coli* strain X82 (*E. coli* Genetic Stock Center, New Haven, CT, USA)) into the strain MBM7007 (W. Dallas, Burroughs Wellcome Co., NC, USA) which carried a deletion of the chromosome from *lac* to *argF*. The *pyrF* mutation confers a pyrimidine requirement on the strain, BA101. In addition, the strain is unable to metabolize cytosine due to the *codAB* deletion. Thus, BA101 is able to grow on minimal medium supplemented with uracil but is unable to utilize cytosine as the sole pyrimidine source.

The construction of BA101 provided a means for positive selection of DNA fragments encoding. The strain, BA101, was transformed with plasmids carrying inserts from the *E. coli* chromosome and the transformants were selected for growth on minimal medium supplemented with cytosine. Using this approach, the transformants were screened for the ability to metabolize cytosine indicating the presence of a DNA fragment encoding CD. Several sources of DNA could be used for the cloning of the *codA* gene: 1) a library of the *E. coli* chromosome could be purchased commercially (for example from Clontech, Palo Alto, CA, USA or Stratagene, La Jolla, CA, USA) and screened; 2) chromosomal DNA could be isolated from *E. coli*, digested with various restriction enzymes and ligated and plasmid

DNA with compatible ends before screening; and/or 3) bacteriophage lambda clones containing mapped *E. coli* chromosomal DNA inserts could be screened.

5 Bacteriophage lambda clones (Y. Kohara, National
Institute of Genetics, Japan) containing DNA inserts
spanning the 6-8 minute region of the *E. coli* chromosome
were screened for the ability to provide transient
complementation of the *codA* defect. Two clones, 137 and
10 138 were identified in this manner. Large-scale
preparations of DNA from these clones were isolated from
500 ml cultures. Restriction enzymes were used to generate
DNA fragments ranging in size from 10-12 kilobases. The
enzymes used were *EcoRI*, *EcoRI* and *BamHI*, and *EcoRI* and
15 *HindIII*. DNA fragments of the desired size were isolated
from preparative agarose gels by electroelution. The
isolated fragments were ligated to pBR322 (Gibco BRL,
Gaithersburg, MD, USA) with compatible ends. The resulting
ligation reactions were used to transform the *E. coli*
20 strain, DH5 α (Gibco BRL, Gaithersburg, MD, USA). This step
was used to amplify the recombinant plasmids resulting from
the ligation reactions. The plasmid DNA preparations
isolated from the ampicillin-resistant DH5 α transformants
were digested with the appropriate restriction enzymes to
25 verify the presence of insert DNA. The isolated plasmid
DNA was used to transform BA101. The transformed cells
were selected for resistance to ampicillin and for the
ability to metabolize cytosine. Two clones were isolated
pEA001 and pEA002. The plasmid pEA001 contains an
30 approximately 10.8 kb *EcoRI*-*BamHI* insert while pEA002
contains an approximately 11.5 kb *EcoRI*-*HindIII* insert.
The isolated plasmids were used to transform BA101 to
ensure that the ability to metabolize cytosine was the
result of the plasmid and not due to a spontaneous
35 chromosomal mutation.

A physical map of the pEA001 DNA insert was generated

using restriction enzymes. Deletion derivatives of pEA001 were constructed based on this restriction map. The resulting plasmids were screened for the ability to allow BA101 to metabolize cytosine. Using this approach, the *codA* gene was localized to a 4.8 kb *EcoRI-BglIII* fragment. The presence of *codA* within these inserts was verified by enzymatic assays for CD activity. In addition, cell extracts prepared for enzymatic assay were also examined by polyacrylamide gel electrophoresis. Cell extracts that were positive for enzymatic activity also had a protein band migrating with an apparent molecular weight of 52,000.

The DNA sequence of both strands was determined for a 1634 bp fragment. The sequence determination began at the *PstI* site and extended to *PvuII* site thus including the *codA* coding domain. An open reading frame of 1283 nucleotides was identified. The thirty amino terminal amino acids were confirmed by protein sequencing. Additional internal amino acid sequences were generated from CNBr-digestion of gel-purified CD.

A 200 bp *PstI* fragment was isolated that spanned the translational start codon of *codA*. This fragment was cloned into pBS⁺. Single-stranded DNA was isolated from 30 ml culture and mutanized using a custom oligonucleotide BA22 purchased from Synthecell Inc., Rockville, MD, USA and the oligonucleotide-directed mutagenesis kit (Amersham, Arlington Heights, IL, USA). The base changes result in the introduction of an *HindIII* restriction enzyme site for joining of CD with CEA TRS and in a translational start codon of ATG rather than GTG. The resulting 90 bp *HindIII-PstI* fragment is isolated and ligated with the remainder of the cytosine deaminase gene. The chimeric CEA TRS/cytosine deaminase gene is created by ligating the *HindIII-PvuII* cytosine deaminase-containing DNA fragment with the CEA TRS sequences.

The strain BA101 and the plasmids, pEA001 and pEA003, were deposited with ATCC under the Budapest Treaty with Accession Nos. 55299, 62916; and 62915 respectively.

5 C) Construction of transcriptional regulatory sequence of
 carcinoembryonic antigen/cytosine deaminase molecular
 chimera

10 A 1508 bp HindIII/PvuII fragment containing the coding
 sequence for cytosine deaminase is isolated from the
 plasmid containing the full length CD gene of Example 13
 that has been altered to contain a HindIII restriction site
 just 5' of the initiation codon. Plasmid pCR92 contains
 CEA sequences -299 to +69 immediately 5' to a unique
15 HindIII restriction site and a polyadenylation signal 3' to
 a unique ApaI restriction site (Example 1A, Figure 4).
 pCR92 is linearised with ApaI, the ends are blunted using
 dNTPs and T4 DNA polymerase, and subsequently digested with
 HindIII. The pCR92 HindIII/ApaI fragment is ligated to the
20 1508 bp HindIII/PvuII fragment containing cytosine
 deaminase. Plasmid pCEA-1/codA, containing CD inserted in
 the appropriate orientation relative to the CEA TRS and
 polyadenylation signal is identified by restriction enzyme
 and DNA sequence analysis.

25

 The optimized CEA-specific TRS, the coding sequence
 for CD with an ATG translation start, and a suitable
 polyadenylation signal are joined together using standard
 molecular biology techniques. The resulting plasmid,
30 containing CD inserted in the appropriate orientation
 relative to the optimized CEA specific TRS and a
 polyadenylation signal is identified by restriction enzyme
 and DNA sequence analysis.

35

Example 2

Construction of a retroviral shuttle vector construct

containing the molecular chimera of Example 1

The retroviral shuttle vector pL-CEA-1/codA is constructed by ligating a suitable restriction fragment containing the optimized CEA TRS/codA molecular chimera including the polyadenylation signal into an appropriate retroviral shuttle vector, such as N2(XM5) linearised at the XhoI site, using standard molecular biology techniques. The retroviral shuttle vector pL-CEA-1/codA is characterized by restriction endonuclease mapping and partial DNA sequencing.

Example 3Virus Production of Retroviral Constructs of Example 3

The retroviral shuttle construct described in Example 2 is placed into an appropriate packaging cell line, such as PA317, by electroporation or infection. Drug resistant colonies, such as those resistant to G418 when using shuttle vectors containing the NEO gene, are single cell cloned by the limiting dilution method, analyzed by Southern blots, and titred in NIH 3T3 cells to identify the highest producer of full-length virus.

Example 4Further data on the CEA TRS

In addition to the plasmids shown in figure 5B, the following combinations of regions have proved particularly advantageous at high level expression of the reporter gene in the system described in Example 1A:

PCR177:

(-14.5kb to -10.6kb) + (-6.1kb to -3.9kb) + (-299b to +69b)

PCR176:

(-13.6kb to -10.6kb) + (-6.1kb to -3.9kb) + (-299b to +69b)

PCR165:

(-3.9kb to -6.1kb) + (4x -90b to +69b)

PCR168:

(-13.6kb to -10.6kb) + (4x -90b to +69b).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: The Wellcome Foundation Limited
(B) STREET: Unicorn House, 160 Euston Road
(C) CITY: London
(E) COUNTRY: G.B.
(F) POSTAL CODE (ZIP): NW1 2BP

(ii) TITLE OF INVENTION: Transcriptional Regulatory Sequence

(iii) NUMBER OF SEQUENCES: 5

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11288 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AAGCTTAAAA CCCAATGGAT TGACACATC AAGAGTTGGA ACAAGTGGAC ATGGAGATGT	60
TACTTGTGGA AATTTAGATG TGTTAGCTA TCGGGCAGGA GAATCTGTGT CAAATTCAG	120
CATGGTTCAG AAGAATCAA AAGTGTACA GTCCAAATGT GCAACAGTGC AGGGGATAAA	180
ACTGTGGTGC ATTCAAACCTG AGGGATATTT TGAACATGA GAAAGGAAGG GATTGCTGCT	240
GCACAGAACA TGGATGATCT CACACATAGA GTTGAAAGAA AGGAGTCAAT CGCAGAATAG	300
AAAATGATCA CTAATTCAC CTCTATAAG TTTCCAAGAG GAAAACCCAA TTCTGCTGCT	360
AGAGATCAGA ATGGAGGTGA CCTGTGCCTT GCAATGGCTG TGAGGGTCAC GGGAGTGTC	420
CTTAGTGCAG GCAATGTGCC GTATCTTAAT CTGGGCAGGG CTTTCATGAG CACATAGGAA	480
TGCAGACATT ACTGCTGTGT TCATTTTACT TCACCGGAAA AGAAGAATAA AATCAGCCGG	540
GCGCGGTGGC TCACGCCTGT AATCCAGCA CTTTGAAGG CTGAGGTGGG CAGATTACTT	600
GAGGTCAGGA GTTCAAGACC ACCCTGGCCA ATATGGTGAA ACCCCGGCTC TACTAAAAAT	660
ACAAAAATTA GCTGGGCATG GTGGTGCGCG CCTGTAATCC CAGCTACTCG GGAGGCTGAG	720

GCTGACAAAT TGCTTGSACC CAGGAAGCAG AGGTTGCAGT GAGCCAAGAT TGTGCCACTG	780
CACTCCAGCT TGGGCAACAG AGCCAGACTC TGTAAAAAA AAAAAAAAAA AAAAAAAAAAG	840
AAAGAAAGAA AAAGAAAGA AAGTATAAAA TCTCTTTGGG TTAACAAAAA AAGATCCACA	900
AAACAAACAC CAGCTCTTAT CAAACTTACA CACTCTGCC AGAGAACAGG AAACACAAAT	960
ACTCATTAACT TCACCTTTGT GGCATAAAAA CTTTCATGTC AAAAGGAGAC CAGGACACAA	1020
TGAGGAAGTA AAACAGCAGG CCTACTTGG GTGCAGAGAG GGAATATCCA CAAATAAAAC	1080
ATTACCAGAA GGAGCTAAGA TTTACTGCAT TGAGTTCATT CCCCAGGTAT GCAAGGTGAT	1140
TTTAACACCT GAAATCAAT CATTCCTTT ACTACATAGA CAGATTAGCT AGAAAAAAT	1200
TACAACTAGC AGAACAGAAG CAATTGGCC TTCTAAAAT TCCACATCAT ATCATCATGA	1260
TGGAGACAGT GCAGACGCCA ATGACAATAA AAAGAGGGAC CTCCTCACC CGGTAAACAT	1320
GTCCACACAG CTCCAGCAAG CACCCGTCTT CCGAGTGAAT CACTGTAACC TCCCCTTTAA	1380
TCAGCCCCAG GCAAGGCTGC CTGCGATGGC CACACAGGCT CCAACCCGTG GGCCTCACC	1440
TCCCGCAGAG GCTCTCCTTT GGCACCCCCA TGGGAGAGC ATGAGGACAG GGCAGAGCCC	1500
TCTGATGCCC ACACATGGCA GGAGCTGAGC CCAGAGCCAT GGGGCTGGA GAGCAGAGCT	1560
GCTGGGGTCA GAGCTTCCTG AGGACACCCA GGCCTAAGGG AAGGCAGCTC CCTGGATGGG	1620
GGCAACCAGG CTCCGGGCTC CAACCTCAGA GCGCGATGG GAGGAGCCAG CACTCTAGGC	1680
CTTTCCTAGG GTGACTCTGA GGGGACCCTG ACACGACAGG ATCGCTGAAT GCACCCGAGA	1740
TGAAGGGGCC ACCACGGGAC CCGTCTCTCG TGGCAGATCA GGAGAGAGTG GGACACCATG	1800
CCAGGCCCCC ATGGCATGGC TGGACTGAC CAGGGCCACT CCCCTGCATG CATCAGCCTC	1860
GGTAAGTCAC ATGACCAAGC CCAGGACCAA TGTGAAGGA AGGAAACAGC ATCCCCTTTA	1920
GTGATGGAAC CCAAGGTCAG TGCAAAGAGA GGCCATGAGC AGTTAGGAAG GGTGGTCCAA	1980
CCTACAGCAC AAACCATCAT CTATCATAAG TAGAAGCCCT GCTCCATGAC CCCTGCATTT	2040
AAATAAACGT TTGTTAAATG AGTCAAATTC CCTCACCATG AGAGCTCACC TGTGTGTAGG	2100
CCCATCACAC ACACAAACAC ACACACACAC ACACACACAC ACACACACAC ACACAGGGAA	2160
AGTGCAGGAT CCTGGACAGC ACCAGGCAGG CTTACAGGC AGAGCAAACA GCGTGAATGA	2220
CCCATGCAGT GCCCTGGGCC CCATCAGCTC AGAGACCCTG TGAGGGCTGA GATGGGGCTA	2280
GGCAGGGGAG AGACTTAGAG AGGGTGGGGC CTCCAGGGAG GGGGCTGCAG GGAGCTGGGT	2340
ACTGCCCTCC AGGGAGGGGG CTGCAGGGAG CTGGGTACTG CCCTCCAGGG AGGGGGCTGC	2400
AGGGAGCTGG GTACTGCCCT CCAGGGAGGG GGCTGCAGGG AGCTGGGTAC TGCCCTCCAG	2460
GGAGGGGGCT GCAGGGAGCT GGGTACTGCC CTCCAGGGAG GCAGGAGCAC TGTTCCTAAC	2520
AGAGAGCACA TCTTCCTGCA GCAGCTGCAC AGACACAGGA GCCCCATGA CTGCCCTGGG	2580
CCAGGGTGTG GATTCCAAAT TTCGTGCCCC ATTGGGTGGG ACGGAGGTTG ACCGTGACAT	2640
CCAAGGGGCA TCTGTGATTC CAACTTAAA CTACTGTGCC TACAAAATAG GAAATAACCC	2700
TACTTTTTCT ACTATCTCAA ATCCCTAAG CACAAGCTAG CACCCTTTAA ATCAGGAAGT	2760

TCAGTCACTC CTGGGGTCCCT CCCATGCCCC CAGTCTGACT TGCAGGTGCA CAGGGTGGCT	2820
GACATCTGTC CTTGCTCCTC CTCTTGGCTC AACTGCCGCC CCTCCTGGGG GTGACTGATG	2880
GTCAGGACAA GGGATCCTAG AGCTGGCCCC ATGATTGACA GGAAGGCAGG ACTTGGCCTC	2940
CATTCTGAAG ACTAGGGGTG TCAAGAGAGC TGGGCATCCC ACAGAGCTGC ACAAGATGAC	3000
GCGGACAGAG GGTGACACAG GGCTCAGGGC TTCAGACGGG TCGGGAGGCT CAGCTGAGAG	3060
TTCAGGGACA GACCTGAGGA GCCTCAGTGG GAAAAGAAGC ACTGAAGTGG GAAGTTCTGG	3120
AATGTTCTGG ACAAGCCTGA GTGCTCTAAG GAAATGCTCC CACCCCGATG TAGCCTGCAG	3180
CACTGGACGG TCTGTGTACC TCCCCGCTGC CCATCCTCTC ACAGCCCCCG CCTCTAGGGA	3240
CACAACTCCT GCCCTAACAT GCATCTTTC TGTCTCATTC CACACAAAAG GGCCTCTGGG	3300
GTCCCTGTTT TGCATTGCAA GGAGTGGAGG TCACGTTCCC ACAGACCACC CAGCAACAGG	3360
GTCCATGGA GGTGCGGTCA GGAGGATCAC ACCTCCCCCC ATGCCCAGGG GACTGACTCT	3420
GGGGGTGATG GATTGGCCTG GAGGCCACTG GTCCCTCTG TCCCTGAGGG GAATCTGCAC	3480
CCTGGAGGCT GCCACATCCC TCCTGATTCT TTCAGCTGAG GGCCCTCTT GAAATCCCAG	3540
GGAGGACTCA ACCCCCCTG GGAAGGGCCC AGTGTGGACG GTTCCACAGC AGCCCACTA	3600
AGGCCCTTGG ACACAGATCC TGAGTGAGAG AACCTTTAGG GACACAGGTG CACGGCCATG	3660
TCCCCAGTGC CCACACAGAG CAGGGGCATC TGGACCCTGA GTGTGTAGCT CCCGCGACTG	3720
AACCCAGCCC TTCCCCATG ACGTGACCCC TGGGGTGGCT CCAGGTCTCC AGTCCATGCC	3780
ACCAAAATCT CCAGATTGAG GGTCCCTCCCT TGAGTCCCTG ATGCCTGTCC AGGAGCTGCC	3840
CCCTGAGCAA ATCTAGAGTG CAGAGGGCTG GGATTGTGGC AGTAAAAGCA GCCACATTG	3900
TCTCAGGAAG GAAAGGGAGG ACATGAGCTC CAGGAAGGGC GATGGCGTCC TCTAGTGGC	3960
GCCTCTGTT AATGAGCAA AAGGGGCCAG GAGAGTTGAG AGATCAGGGC TGGCCTTGA	4020
CTAAGGCTCA GATGGAGAGG ACTGAGGTGC AAAGAGGGGG CTGAAGTAGG GGAGTGGTCG	4080
GGAGAGATGG GAGGAGCAGG TAAGGGGAAG CCCCAGGGAG GCCGGGGGAG GGTACAGCAG	4140
AGCTCTCCAC TCCTCAGCAT TGACATTGG GGTGGTCGTG CTAGTGGGGT TCTGTAAATT	4200
GTAGGGTGT CAGCACCATC TGGGGACTCT ACCCACTAAA TGCCAGCAGG ACTCCCTCCC	4260
CAAGCTCTAA CAACCAACAA TGTCTCCAGA CTTTCCAAAT GTCCCTGGA GAGCAAAATT	4320
GCTTCTGGCA GAATCACTGA TCTACGTGAG TCTCTAAAAG TGACTCATCA GCGAAATCCT	4380
TCACCTCTTG GGAGAAGAAT CACAAGTGTG AGAGGGGTAG AACTGCAGA CTTCAAAATC	4440
TTTCCAAAAG AGTTTTACTT AATCAGCAGT TTGATGTCCC AGGAGAAGAT ACATTTAGAG	4500
TGTTTAGAGT TGATGCCACA TGGCTGCCTG TACCTCACAG CAGGAGCAGA GTGGGTTTTT	4560
CAAGGGCCTG TAACCACAAC TGGAAATGACA CTCACTGGGT TACATTACAA AGTGGAATGT	4620
GGGGAATTCT GTAGACTTTG GGAAGGGAAA TGTATGACGT GAGCCACAG CCTAAGGCAG	4680
TGGACAGTCC ACTTTGAGGC TCTCACCATC TAGGAGACAT CTCAGCCATG AACATAGCCA	4740
CATCTGTCTAT TAGAAAACAT GTTTTATTAA GAGGAAAAAT CTAGGCTAGA AGTGCTTIAT	4800

GCTCTTTTTT CTCTTTATGT TCAAATTCAT ATACTTTTAG ATCATTCCCTT AAAGAAGAAT	4860
CTATCCCCCT AAGTAAATGT TATCACTGAC TGGATAGTGT TGGTGTCTCA CTCCCACCC	4920
CTGTGTGGTG ACAGTGCCCT GCTTCCCCAG CCGTGGGCCC TCTCTGATTC CTGAGAGCTT	4980
TGGGTGCTCC TTCATTAGGA GGAAGAGAGG AAGGGTGTTC TTAATATTCT CACCATTTC	5040
CCATCCACCT CTTAGACACT GGAAGAATC AGTTGCCAC TCTGGGATTT GATCCTCGAA	5100
TTAATGACCT CTATTTCTGT CCCTTGTCCTA TTTCACAAAT GTGACAGGCC TAAGAGGTGC	5160
CTTCTCCATG TGATTTTTGA GGAGAAGGTT CTCAGATAA GTTTCTCAC ACCTCTTTGA	5220
ATTACCTCCA CCTGTGTCCC CATCACCATT ACCAGCAGCA TTTGGACCTT TTTCTCTTA	5280
GTCAGATGCT TTCCACCTCT TGAGGGTGTA TACTGTATGC TCTCTACACA GGAATATGCA	5340
GAGGAAATAG AAAAAGGGAA ATCGCATTAC TATTCAGAGA GAAGAAGACC TTTATGTGAA	5400
TGAATGAGAG TCTAAAATCC TAAGAGAGCC CATATAAAT TATTACCAGT GCTAAAATA	5460
CAAAAGTTAC ACTAACAGTA AACTAGAATA ATAAACATG CATCACAGTT GCTGGTAAAG	5520
CTAAATCAGA TATTTTTTTC TTAGAAAAAG CATCCCATGT GTGTTGCAGT GATGACAGGA	5580
GTGCCCTTCA GTCAATATGC TGCCTGTAAT TTTTGTCCC TGGCAGAATG TATTGTCTTT	5640
TCTCCCTTTA AATCTTAAAT GCAAACTAA AGGCAGCTCC TGGGCCCCCT CCCCAGTC	5700
AGCTGCCTGC AACCAGCCCC ACGAAGAGCA GAGGCCTGAG CTTCCTGGT CAAAATAGGG	5760
GGCTAGGGAG CTTAACCTTG CTCGATAAAG CTGTGTCCC AGAATGTGCG TCCTGTCCC	5820
AGGGGCACCA GCCTGGAGGG TGCTGAGCCT CACTGGTGGC CTGATGCTTA CCTGTGCCC	5880
TCACACCAGT GGTCACTGGA ACCTTGAAC CTGGCTGTC GCCCGGATCT GCAGATGTCA	5940
AGAACTTCTG GAAGTCAAAT TACTGCCCA TCTCCAGGG CAGATACCTG TGAACATCCA	6000
AAACCATGCC ACAGAACCCT GCCTGGGGTC TACAACACAT ATGGACTGTG AGCACCAAGT	6060
CCAGCCCTGA ATCTGTGACC ACCTGCCAAG ATGCCCTAA CTGGGATCCA CCAATCACTG	6120
CACATGGCAG GCAGCGAGG TTGGAGGTGC TTCGCCCAA GGCAGCCCCA ATTTGCTGGG	6180
AGTTTCTTGG CACCTGGTAG TGGTGAGGAG CCTTGGGACC CTCAGGATTA CTCCCCTAA	6240
GCATAGTGGG GACCCTTCTG CATCCCCAGC AGGTGCCCCG CTCTTCAGAG CCTCTCTCTC	6300
TGAGGTTTAC CCAGACCCCT GCACCAATGA GACCATGCTG AAGCCTCAGA GAGAGAGATG	6360
GAGCTTTGAC CAGGAGCCGC TCTTCCTTGA GGGCCAGGGC AGGGAAGCA GGAGGCAGCA	6420
CCAGGAGTGG GAACACCAGT GTCTAAGCCC CTGATGAGAA CAGGGTGGTC TCTCCCATAT	6480
GCCCATACCA GGCCTGTGAA CAGAATCCTC CTCTGTCAGT GACAATGTCT GAGAGGACGA	6540
CATGTTTCCC AGCCTAACGT GCAGCCATGC CATCTACCC ACTGCCTACT GCAGGACAGC	6600
ACCAACCCAG GAGCTGGGAA GCTGGGAGAA GACATGGAAT ACCCATGGCT TCTCACCTTC	6660
CTCCAGTCCA GTGGGCACCA TTTATGCCTA GGACCCAC CTGCCGGCCC CAGGCTCTTA	6720
AGAGTTAGGT CACCTAGGTG CCTCTGGGAG GCCGAGGCAG GAGAATTGCT TGAACCCGGG	6780
AGGCAGAGGT TGCAGTGAGC CGAGATCACA CCACTGCACT CCAGCCTGGG TGACAGAATG	6840

AGACTCTGTC	TCAAAAAAAAA	AGAGAAAGAT	AGCATCAGTG	GCTACCAAGG	GCTAGGGGCA	6900
GGGGAAAGTG	GAGAGTTAAT	GATTAATAGT	ATGAAGTTTC	TATGTGAGAT	GATGAAAATG	6960
TTCTGGAAAA	AAAAATATAG	TGGTGAGGAT	GTAGAATATT	GTGAATATAA	TTAACGGCAT	7020
TTAATTGTAC	ACTTAACATG	ATTAATGTGG	CATATTTTAT	CTTATGTATT	TGACTACATC	7080
CAAGAAACAC	TGGGAGAGGG	AAAGCCCACC	ATGTAAAATA	CACCCACCCT	AATCAGATAG	7140
TCCTCATTTT	ACCCAGGTAC	AGGCCCCCTCA	TGACCTGCAC	AGGAATAACT	AAGGATTAA	7200
GGACATGAGG	CTTCCCAGCC	AACCTGCAGG	GCACACATA	AATGTATCTG	CAACACAGACT	7260
GAGAGTAAAG	CTGGGGGCAC	AAACCTCAGC	ACTGCCAGGA	CACACACCCT	TCTCTGGAT	7320
TCTGACTTTA	TCTGACCCGG	CCCACTGTCC	AGATCTTGT	GTGGGATTGG	GACAAGGGAG	7380
GTCTAATAGC	CTGTCCCCAG	GGCACTCTGT	GTAGGCACAC	GAGACCTCCC	CACCCCCCA	7440
CCGTTAGGTC	TCCACACATA	GATCTGACCA	TAGGCATTG	TGAGGAGGAC	TCTAGCCGG	7500
GCTCAGGGAT	CACACCAGAG	AATCAGGTAC	AGAGAGGAAG	ACGGGGCTCG	AGGAGCTGAT	7560
GGATGACACA	GAGCAGGTTT	CCTGCAGTCC	ACAGGTCCAG	CTCACCTTGG	TGTAGGTGCC	7620
CCATCCCCCT	GATCCAGGCA	TCCCTGACAC	AGCTCCCTCC	CGGAGCCTCC	TCCCAGGTGA	7680
CACATCAGGG	TCCCTCACTC	AAGCTGTCCA	GAGAGGGCAG	CACCTTGGAC	AGCGCCACC	7740
CCACTTCACT	CTTCTCCCT	CACAGGGCTC	AGGGCTCAGG	GCTCAAGTCT	CAGAACAAT	7800
GGCAGAGGCC	AGTGAGCCCA	GAGATGGTGA	CAGGGCAATG	ATCCAGGGGC	AGCTGCCTGA	7860
AACGGGAGCA	GGTGAAGCCA	CAGATGGGAG	AAGATGGTTC	AGGAAGAAAA	ATCCAGGAAT	7920
GGGCAGGAGA	GGAGAGGAGG	ACACAGGCTC	TGTGGGGCTG	CAGCCCAGGA	TGGGACTAAG	7980
TGTGAAGACA	TCTCAGCAGG	TGAGGCCAGG	TCCCATGAAC	AGAGAAGCAG	CTCCACCTC	8040
CCCTGATGCA	CGGACACACA	GAGTGTGTGG	TGCTGTGCCC	CCAGAGTCGG	GCTCTCCTGT	8100
TCTGGTCCCC	AGGGAGTGAG	AAGTGAGGTT	GACTTGTCCC	TGCTCCTCTC	TGCTACCCCA	8160
ACATTACCTT	TCTCCTCATG	CCCCTCTCTC	TCAATATGA	TTTGGATCTA	TGTCCCCGCC	8220
CAAACTCTCAT	GTCAAATTGT	AAACCCCAAT	GTGGAGGTG	GGGCCTTGTG	AGAAGTGATT	8280
GGATAATGCG	GGTGGATTTT	CTGCTTTGAT	GCTGTTTCTG	TGATAGAGAT	CTCACATGAT	8340
CTGTTTGT	AAAAGTGTGT	AGCACCTCTC	CCCTCTCTCT	CTCTCTCTCT	TACTCATGCT	8400
CTGCCAIGTA	AGACGTTTCT	GTTTCCCTTT	CACCGTCCAG	AATGATTGTA	AGTTTCTGA	8460
GGCCTCCCCA	GGAGCAGAAG	CCACTATGCT	TCTGTACAA	CTGCAGAATG	ATGAGCGAAT	8520
TAAACCTCTT	TTCTTTATAA	ATTACCCAGT	CTCAGGTATT	TCTTTATAGC	AATGCGAGGA	8580
CAGACTAATA	CAATCTTCTA	CTCCCAGATC	CCGSCACAGC	CTTAGCCCCA	GACATCACTG	8640
CCCCTGGGAG	CATGCACAGC	GCAGCCTCCT	GCCGACAAAA	GCAAAGTCAC	AAAAGGTGAC	8700
AAAAATCTGC	ATTTGGGGAC	ATCTGATTGT	GAAAGAGGGA	GGACAGTACA	CTTGTAGCCA	8760
CAGAGACTGG	GGCTCACCGA	GCTGAAACCT	GTAGCACTT	TGGCATAACA	TGTGCATGAC	8820
CCGTGTTCAA	TGTCTAGAGA	TCAGTGTGA	GTAACACAGC	CTGGTCTGGG	GCCGCTGCTG	8880

TCCCCACTTC CCTCCTGTCC ACCAGAGGGC GGCAGAGTTC CTCCCACCCT GGAGGCTCCC	8940
CAGGGGCTGC TGACCTCCCT CAGCCGGGCC CACAGCCCAG CAGGGTCCAC CCTCACCCGG	9000
GTCACCTCGG CCCACGTCCCT CCTCGCCCTC CGAGCTCCTC ACACGGACTC TGTGAGCTCC	9060
TCCCTGCAGC CTATCGGCCG CCCACCTGAG GOTTGTGGC CGCCCACTTG AGGCCTGTG	9120
GCTGCCCTCT GCAGGCAGGT CCTGTCCCCT ACACCCCTC CTTCGCCGGG CTCAGCTGA	9180
AGGGCGTCTC CCAGGGCAGC TCCCTGTGAT TCCAGGACA GCTCAGTCTC TCACAGGCTC	9240
CGACGCCCCC TATGCTGTCA CCTCACAGCC CTGTATTAC CATTAAGTCC TCAGTCCAT	9300
GAAGTTCATC GAGCGCTGT CTCCCGTTA CAGGAAACT CTGTGACAGG GACCACTCT	9360
GTCTGTCTCT CTGTGGAATC CCAGGGCCCA GCGCAGTGC TGACACGGAA CAGATGTCC	9420
ATAAATACTG GTTAAATGTG TGGGAGATCT CTTAAAGAA GCATATCACC TCCGTGTGGC	9480
CCCCAGCAGT CAGAGTCTGT TCCATGTGGA CACAGGGGCA CTGGCACCAG CATGGGAGGA	9540
GGCCAGCAAG TGCCCGCGGC TGCCCCAGGA ATGAGGCCTC AACCCCAAGA GCTTCAGAG	9600
GGAGGACAGA GGCCTGCAGG GAATAGATCC TCCGGCCTGA CCCTGCAGCC TAATCCAGG	9660
TTCAGGGTCA GCTCACACCA CGTCGACCCT GGTGACATC CCTAGGGCAG TTCCAGACAA	9720
GGCCGGAGGT CTCTCTTGC CCTCCAGGGG GTGACATTGC ACACAGACAT CACTCAGGAA	9780
ACGGATTCCC CTGGACAGGA ACCTGGCTTT GCTAAGGAAG TGGAGGTGGA GCCTGTTTC	9840
CATCCCTTGC TCCAACAGAC CCTTCTGATC TCTCCACAT ACCTGCTCTG TTCCTTCTG	9900
GGTCCTATGA GGACCCTGTT CTGCCAGGGG TCCCTGTGCA ACTCCAGACT CCCTCCTGGT	9960
ACCACCATGG GGAAGGTGGG GTGATCACAG GACAGTCAGC CTCGCAGAG CAGAGACCAC	10020
CCAGGACTGT CAGGGAGAAC ATGGACAGGC CCTGAGCCGC AGCTCAGCA ACAGACACGG	10080
AGAGGGAGGG TCCCCCTGGA GCCTTCCCCA AGGACAGCAG AGCCAGAGT CACCCACCTC	10140
CCTCCACCAC AGTCCTCTCT TTCCAGGACA CACAAGACAC CTCCCCCTCC ACATGCAGGA	10200
TCTGGGACT CCTGAGACCT CTGGGCCTGG GTCTCCATCC CTGGGTCACT GCGGGGTTG	10260
GTGGTACTGG AGACAGAGGG CTGGTCCCTC CCCAGCCACC ACCCAGTGAG CCTTTTCTA	10320
GCCCCCAGAG CCACCTCTGT CACCTTCTG TTGGGCATCA TCCCACCTTC CCAGAGCCCT	10380
GGAGAGCATG GGGAGACCCG GGACCCTGCT GGGTTTCTCT GTCACAAAGG AAAATAATCC	10440
CCCTGGTGTG ACAGACCCAA GGACAGAACA CAGCAGAGGT CAGCACTGGG GAAGACAGGT	10500
TGTCCTCCCA GGGGATGGGG GTCCATCCAC CTGCGGAAA AGATTGTCT GAGGAACTGA	10560
AAATAGAAGG GAAAAAGAG GAGGGACAAA AGAGGCAGAA ATGAGAGGGG AGGGGACAGA	10620
GGACACCTGA ATAAAGACCA CACCCATGAC CCACGTGATG CTGAGAAGTA CTCCTGCCCT	10680
AGGAAGAGAC TCAGGGCAGA GGGAGGAAGG ACAGCAGACC AGACAGTCAC AGCAGCTTG	10740
ACAAAACGTT CCTGGAATC AAGCTCTTCT CCACAGAGGA GGACAGAGCA GACAGCAGAG	10800
ACCATGGAGT CTCCCTCGGC CCCTCCCCAC AGATGGTGCA TCCCTGGCA GAGGCTCTG	10860
CTCACAGGTG AAGGGAGGAC AACCTGGGAG AGGGTGGGAG GAGGGAGCTG GGGTCTCTG	10920

GGTAGGACAG GGCTGTGAGA CGGACAGAGG GCTCCTGTTG GAGCCTGAAT AGGGAAGAGG	10980
ACATCAGAGA GGGACAGGAG TCACACCAGA AAAATCAAAT TGAAGTGGAA TTGGAAGGG	11040
GCAGGAAAC CTCAGAGTT CTATTTTCCT AGTTAATTGT CACTGGCCAC TACGTTTTTA	11100
AAAATCATAA TAACTGCATC AGATGACACT TTAATAAAA ACATAACCAG GGCATGAAC	11160
ACTGTCCTCA TCCGCCTACC GCGGACATTG GAAAATAAGC CCCAGGCTGT GGAGGGCCCT	11220
GGGAACCCCTC ATGAACTCAT CCACAGGAAT CTGCAGCCTG TCCCAGGCAC TGGGGTGCAA	11280
CCAAGATC	11288

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3774 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

AAGCTTTTTA GTGCTTTAGA CAGTGAGCTG GTCTGTCTAA CCCAAGTGAC CTGGGCTCCA	60
TACTCAGCCC CAGAAGTGAA GGGTGAAGCT GGGTGGAGCC AAACCAGGCA AGCCTACCCT	120
CAGGGCTCCC AGTGGCCTGA GAACCATGG ACCCAGGACC CATTACTTCT AGGGTAAGGA	180
AGGTACAAAC ACCAGATCCA ACCATGGTCT GGGGGGACAG CTGTCAAATG CCTAAAAATA	240
TACCTGGGAG AGGAGCAGGC AACTATCAC TGCCCCAGGT TCTCTGAACA GAAACAGAGG	300
GGCAACCCAA AGTCCAAATC CAGGTGAGCA GGTGCACCAA ATGCCCAGAG ATATGACCAG	360
GCAAGAAGTG AAGGAACCAAC CCCTGCATCA AATGTTTTGC ATGGGAAGGA GAAGGGGGTT	420
GCTCATGTTT CCAATCCAGG AGAATGCATT TGGGATCTGC CTTCTTCTCA CTCCTTGTTT	480
AGCAAGACTA AGCAACCAGG ACTCTGGATT TGGGGAAAGA CGTTTATTTG TGGAGGCCAG	540
TGATGACAAT CCCACGAGGG CCTAGGTGAA GAGGGCAGGA AGGCTCGAGA CACTGGGGAC	600
TGAGTGAAAA CCACACCCAT GATCTGCACC ACCCATGGAT GCTCCTTCAT TGCTCACCTT	660
TCTGTTGATA TCAGATGGCC CCATTTTCTG TACCTTCACA GAAGGACACA GGCTAGGGTC	720
TGTGCATGGC CTTTCATCCC GGGGCCATGT GAGGACAGCA GGTGGGAAAG ATCATGGGTC	780
CTCCTGGGTC CTGCAGGGCC AGAACATTCA TCACCCATAC TGACCTCCTA GATGGGAATG	840
GCTTCCCTGG GGCTGGGCCA ACGGGGCCTG GGCAGGGGAG AAAGGACGTC AGGGGACAGG	900
GAGGAAGGGT CATCGAGACC CAGCCTGGAA GGTTCCTGTC TCTGACCATC CAGGATTAC	960
TTCCTGTCAT CTACCTTTGG TCATTTTCCC TCAGCAATGA CCAGCTCTGC TTCCTGATCT	1020

CAGCCTCCCA CCCTGGACAC AGCACCCCAG TCCCTGGCCC GGCTGCATCC ACCCAATACC	1080
CTGATAACCC AGGACCCATT ACTTCTAGGG TTAGGAGGGT CCAGGAGACA GAAGCTGAGG	1140
AAAGGTCTGA AGAAGTCACA TCTGTCCTGG CCAGAGGGGA AAAACCATCA GATGCTGAAC	1200
CAGGAGAATC TTGACCCAGG AAAGGGACCG AGGACCCAAG AAAGGAGTCA GACCCACCAGG	1260
GTTTGCCTGA GAGGAAGGAT CAAGGCCCCG AGGGAAGCA GGGCTGGCTG CATGTGCAGG	1320
ACACTGGTGG GGCATATGTC TCTTAGATTG TCCCTGAATT GAGTGTCCCT GCCATGGCCA	1380
GACTCTCTAC TCAGGCCTGG ACATGCTGAA ATAGGACAAT GGCCTTGTCC TCTCTCCCCA	1440
CCATTTGGCA AGAGACATTA AGGACATTCC AGGACATGCC TTCCTGGGAG GTCCAGGTTC	1500
TCTGTCTCAC ACCTCAGGSA CTGTAGTTAC TGCATCAGCC ATGGTAGGTG CTGATCTCAC	1560
CCAGCCTGTC CAGGCCCTTC CACTCTCCAC TTTGTGACCA TGTCCAGGAC CACCCCTCAG	1620
ATCCTGAGCC TGCAAATACC CCCTTGCTGG GTGGGTGGAT TCAGTAAACA GTGAGCTCCT	1680
ATCCAGCCCC CAGAGCCACC TCTGTACCT TCCGTCTGGG CATCATCCCA CCTTCACAAG	1740
CACTAAGAG CATGGGGAGA CCTGGCTAGC TGGGTTTCTG CATCACAAG AAAATAATCC	1800
CCCAGGTTCC GATTCCCRGG GCTCTGTATG TGGAGCTGAC AGACCTGAGG CCAGGAGATA	1860
GCAGAGGTCA GCCCTAGGSA GGGTGGGTCA TCCACCCAGG GGACAGGGGT GCACCAGCCT	1920
TGCTACTGAA AGGGCCTCCC CAGGACAGCG CCATCAGCCC TGCCTGAGAG CTTTGCTAAA	1980
CAGCAGTCAG AGGAGGCCAT GGCAGTGGCT GAGCTCCTGC TCCAGGCCCC AACAGACCAG	2040
ACCAACAGCA CAATGCAGTC CTTCCCCAAC GTCACAGGTC ACCAAGGGA AACTGAGGTG	2100
CTACCTAACC TTAGAGCCAT CAGGGGAGAT AACAGCCCAA TTCCCAAAC AGGCCAGTTT	2160
CAATCCCATG ACAATGACCT CTCTGCTCTC ATTCTTCCCA AAATAGGACG CTGATTCTCC	2220
CCCACCATGG ATTTCTCCCT TGTCCCGGGA GCCTTTTCTG CCCCCTATGA TCTGGGCACT	2280
CCTGACACAC ACCTCCTCTC TGGTGACATA TCAGGGTCCC TCACTGTCAA GCAGTCCAGA	2340
AAGGACAGAA CCTTGGACAG CGCCCATCTC AGCTTCACCC TTCCTCCTTC ACAGGGTTCA	2400
GGGCAAAGAA TAAATGGCAG AGGCCAGTGA GCCCAGAGAT GGTGACAGGC AGTGACCCAG	2460
GGGCAGATGC CTGGAGCAGG AGCTGGCGGG GCCACAGGGA GAAGGTGATG CAGGAAGGGA	2520
AACCCAGAAA TGGGCAGGAA AGGAGGACAC AGGCTCTGTG GGGCTGCAGC CCAGGGTTGG	2580
ACTATCAGTG TGAAGCCATC TCAGCAAGTA AGGCCAGGTC CCATGAACAA GAGTGGGAGC	2640
ACGTGGCTTC CTGCTCTGTA TATGGGGTGG GGGATTCCAT GCCCCATAGA ACCAGATGGC	2700
CGGGGTTTCA ATGGAGAAGG AGCAGGACAG GGGATCCCCA GGATAGGAGG ACCCCAGTGT	2760
CCCCACCCAG GCAGGTGACT GATGAATGGG CATGCAGGGT CCTCCTGGGC TGGGCTCTCC	2820
CTTTGTCCCT CAGGATTCCCT TGAAGGAACA TCCGGAAGCC GACCACATCT ACCTGGTGGG	2880
TTCTGGGGAG TCCATGTAAA GCCAGGAGCT TGTGTTGCTA GGAGGGGTCA TGGCATGTGC	2940
TGGGGGCACC AAAGAGAGAA ACCTGAGGGC AGGCAGGACC TGGTCTGAGG AGGCATGGGA	3000
GCCCAGATGG GGAGATGGAT GTCAGGAAAG GCTCCCCCAT CAGGGAGGGT GATAGCAATG	3060

39

```

GGGGGTCTGT GGGAGTGGGC ACGTGGGATT CCCTGGGCTC TGCCAAGTTC CCTCCCATAG 3120
TCACAACCTG GGGACACTGC CCATGAAGGG GCGCCTTTGC CCAGCCAGAT GCTGCTGGTT 3180
CTGCCCCATCC ACTACCTCTT CTGCTCCAGC CACTCTGGGT CTTTCTCCAG ATGCCCTGGA 3240
CAGCCCTGGC CTGGGCCTGT CCCCTGAGAG GTGTTGGGAG AAGCTGAGTC TCTGGGGACA 3300
CTCTCATCAG AGTCTGAAG GCACATCAGG AAACATCCCT GGTCTCCAGG ACTAGGCAAT 3360
GAGGAAAGGG CCCCAGCTCC TCCCTTTGCC ACTGAGAGGG TCGACCTGG GTGGCCACAG 3420
TGACTTCTGC GTCTGTCCCA GTCACCTGA AACCACAACA AAACCCAGC CCCAGACCTT 3480
GCAGGTACAA TACATGTGG GACAGTCTGT ACCCAGGGGA AGCCAGTTCT CTCTTCCTAG 3540
GAGACCGGGC CTCAGGGCTG TGCCCGGGGC AGGCGGGGGC AGCAGTGCC TGTCCITGAG 3600
AACTCGGGAC CTTAAGGGTC TCTGCTCTGT GAGGCACAGC AAGGATCCTT CTGTCCAGAG 3660
ATGAAAGCAG CTCCTGCCCC TCCTCTGACC TCTTCTCCT TCCCAAATCT CAACCAACAA 3720
ATAGGTGTTT CAAATCTCT CATCAAATCT TCATCCATCC ACATGAGAAA GCTT 3774

```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
CCCTGTGATC TCCAGGACAG CTCAGTCTCC GTCCAATCTC 40
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
GTTCCTGAG TGATGCTGT GTGCAATG
```

28

40

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCTGGAAGCTC AAGCTTGAAT TCTCCACAGA GGAGG

35

CLAIMS:

1. A DNA molecule comprising the carcinoembryonic antigen (CEA) transcriptional regulatory sequence (TRS) but without associated CEA coding sequence.
2. A molecular chimaera comprising a CEA TRS and a DNA sequence operatively linked thereto encoding a heterologous enzyme.
3. A molecular chimaera according to claim 2 wherein the heterologous enzyme is capable of catalysing the production of an agent cytotoxic or cytostatic to CEA⁺ cells.
4. A molecular chimaera according to claim 3 wherein the heterologous enzyme is cytosine deaminase (CD).
5. A molecular chimaera according to any of claims 2 to 4 wherein the CEA TRS and the sequence encoding a heterologous enzyme are in an expression cassette.
6. A molecular chimaera according to claim 5 which comprises DNA sequence of the coding sequence of the gene coding for the heterologous enzyme and additionally includes an appropriate polyadenylation sequence which is linked downstream in a 3' position and in proper orientation to the CEA TRS.
7. A retroviral shuttle vector comprising a molecular chimaera according to any of claims 2 to 6.
8. A retroviral shuttle vector according to claim 7 comprising a DNA sequence comprising a 5' viral LTR sequence, a cis acting psi encapsidation sequence, the molecular chimaera and a 3' viral LTR sequence.

9. A retroviral shuttle vector according to claim 8 based on Moloney murine leukaemia virus.

10. A retroviral shuttle vector according to any of claims 7 to 9 which is a SIN vector.

11. An infective virion comprising a retroviral shuttle vector according to any of claims 7 to 10, the vector being encapsidated within viral proteins to create an artificial, infective, replication defective, retrovirus.

12. A packaging cell line comprising a retroviral shuttle vector according to any of claims 7 to 10.

13. A pharmaceutical composition comprising an infective virion according to claim 11 or packaging cell line according to claim 12 together with a pharmaceutically acceptable carrier.

14. Use of CEA TRS for targeting expression of a heterologous enzyme to CEA⁺ cells.

15. Use according to claim 14 wherein the heterologous enzyme is capable of catalysing the production of an agent cytotoxic or cytostatic to CEA⁺ cells.

16. Use according to claim 15 wherein the heterologous enzyme is CD.

17. A DNA molecule according to claim 1 which comprises one or more of the following sequence regions of the CEA gene in either orientation:

about -299b to about +69b, more preferably about -90b to about +69b;

-14.4kb to -10.6kb, preferably -13.6kb to -10.6kb;

-6.1kb to -3.8kb.

18. A molecular chimaera according to any of claims 2 to 6, retroviral shuttle vector according to any claims 7 to 10, packaging cell line according to claim 12 or composition according to claim 13 wherein the CEA TRS comprises one or more of the following sequence regions of the CEA gene in either orientation:

about -299b to about +69b, more preferably about -90b to about +60b;

-14.4kb to -10.6kb, preferably -13.6kb to -10.6kb;

-6.1kb to -3.8kb.

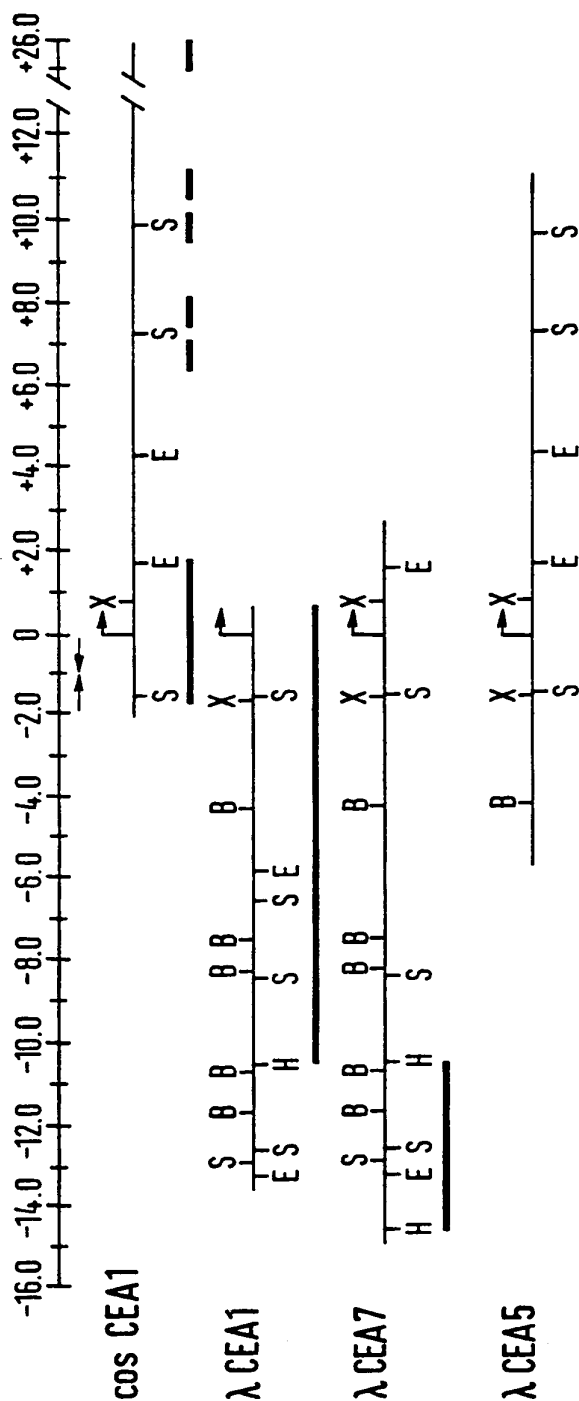
19. Use according to any of claims 14 to 16 wherein the CEA TRS comprises one or more of the following sequence regions of the CEA gene in either orientation:

about -199b to about +69b, more preferably about -90b to about +69b;

-14.4kb to -10.6kb, preferably -13.6kb to -10.6kb;

-6.1kb to -3.8kb.

1/20



cos CEA1 MAP AND SEQUENCE FROM SCHREWE, et.al. Mol Cell Biol 10:2738, 1990

Fig. 1

2/20

λCEA1

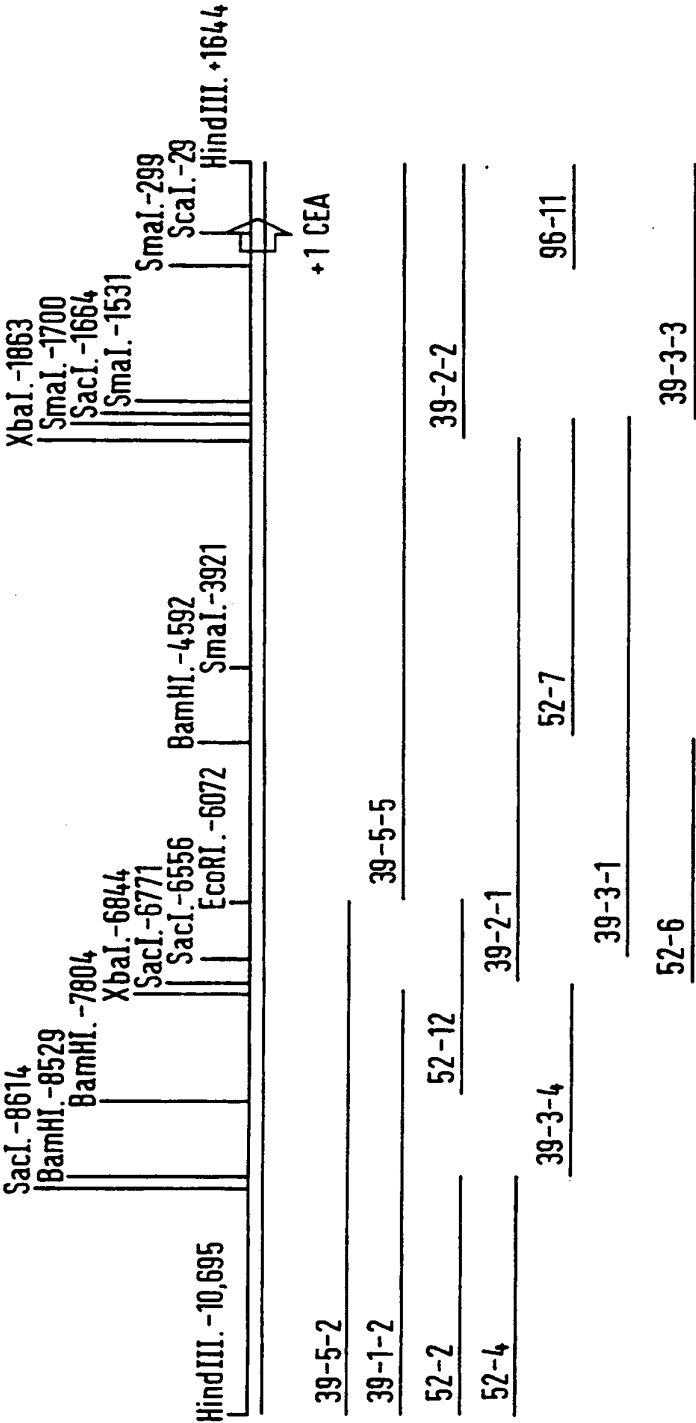


Fig. 2

3/20

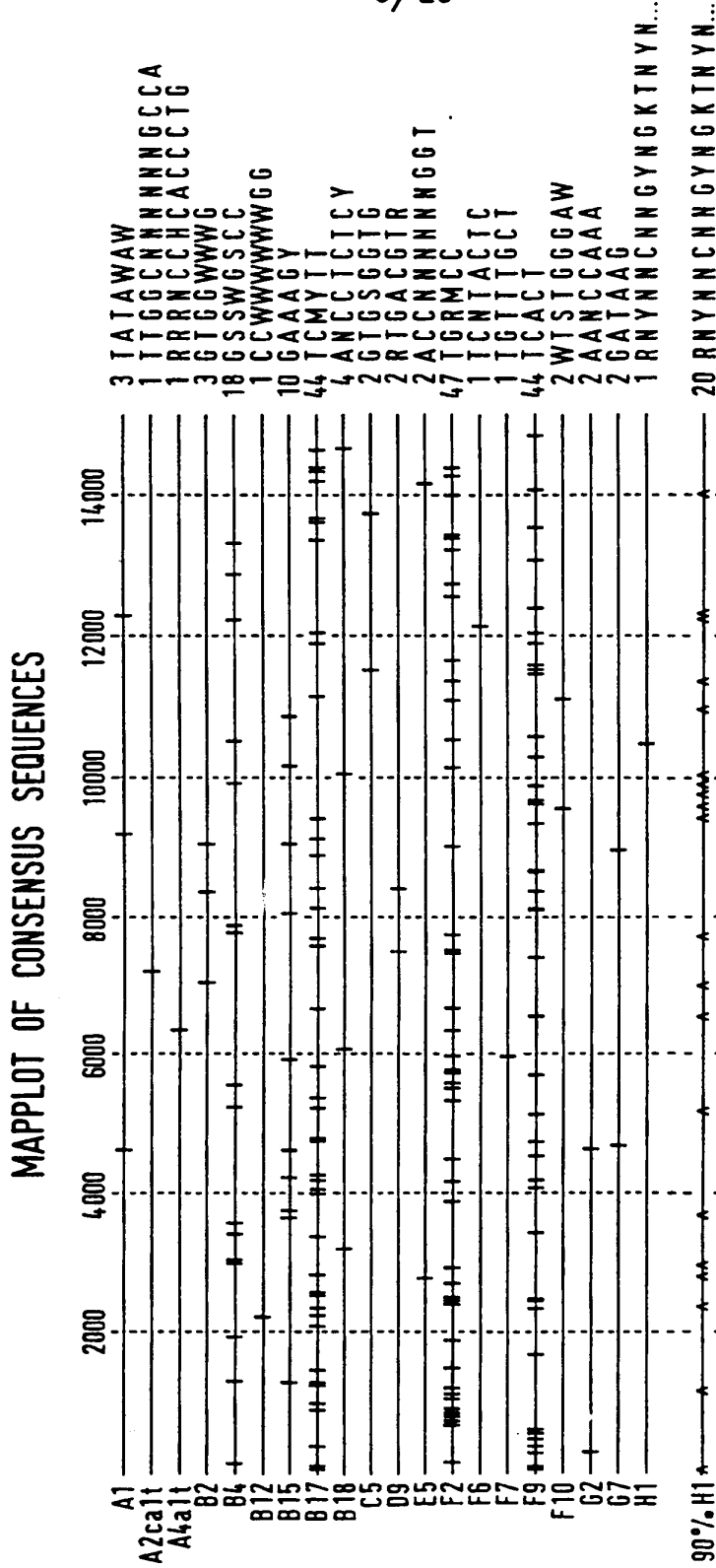


Fig. 3

4/20

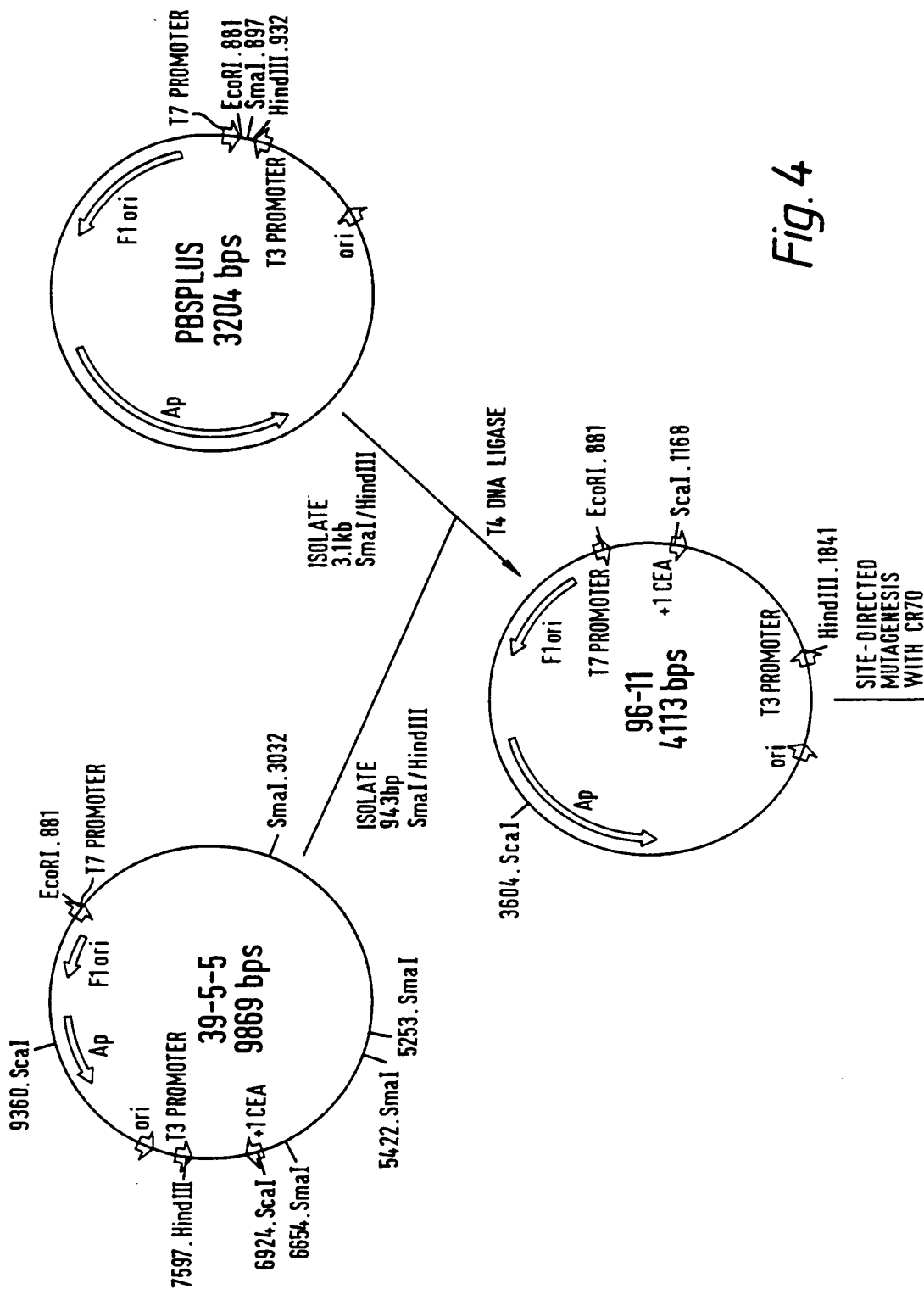


Fig. 4

5/20

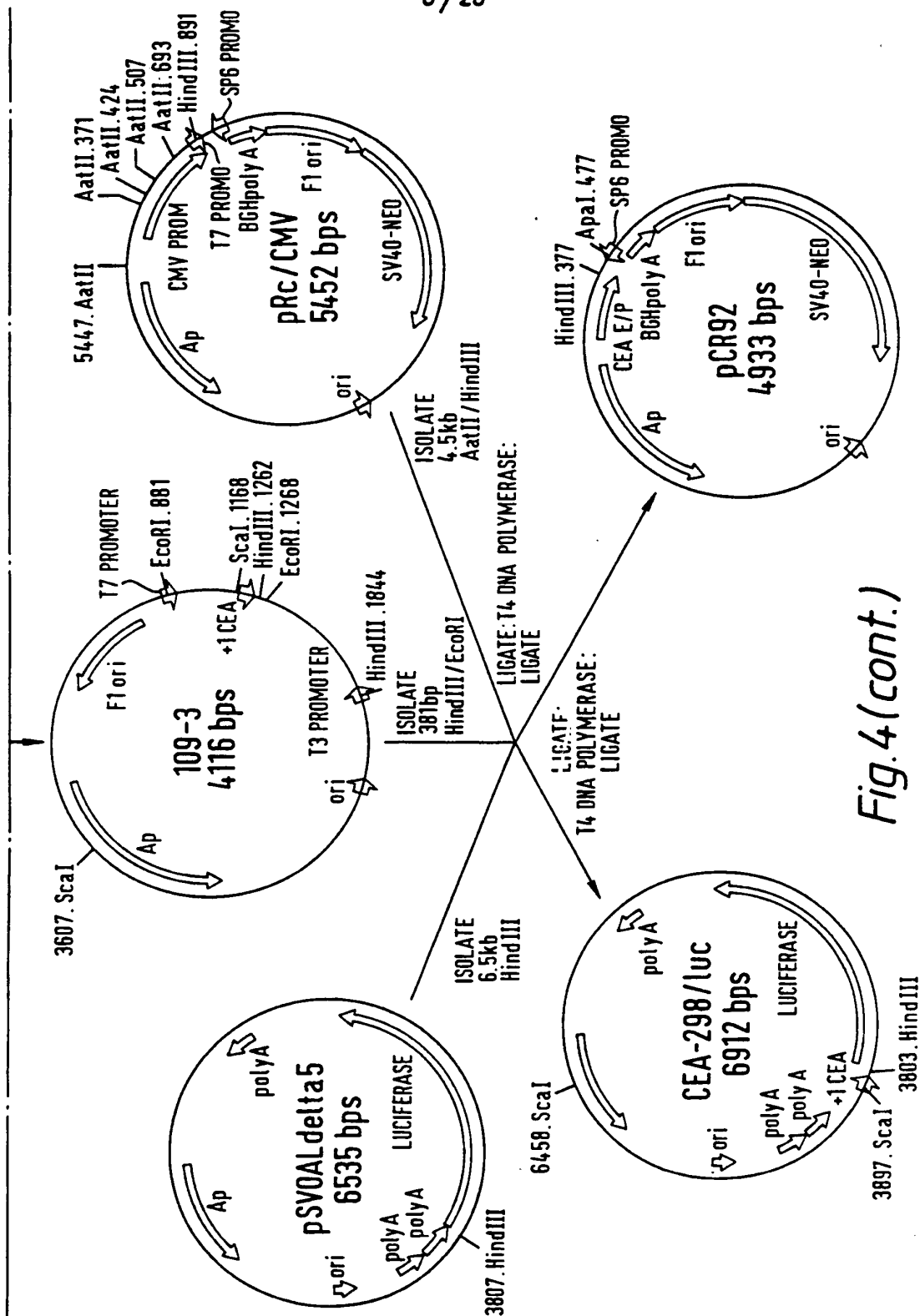


Fig. 4(cont.)

6/20

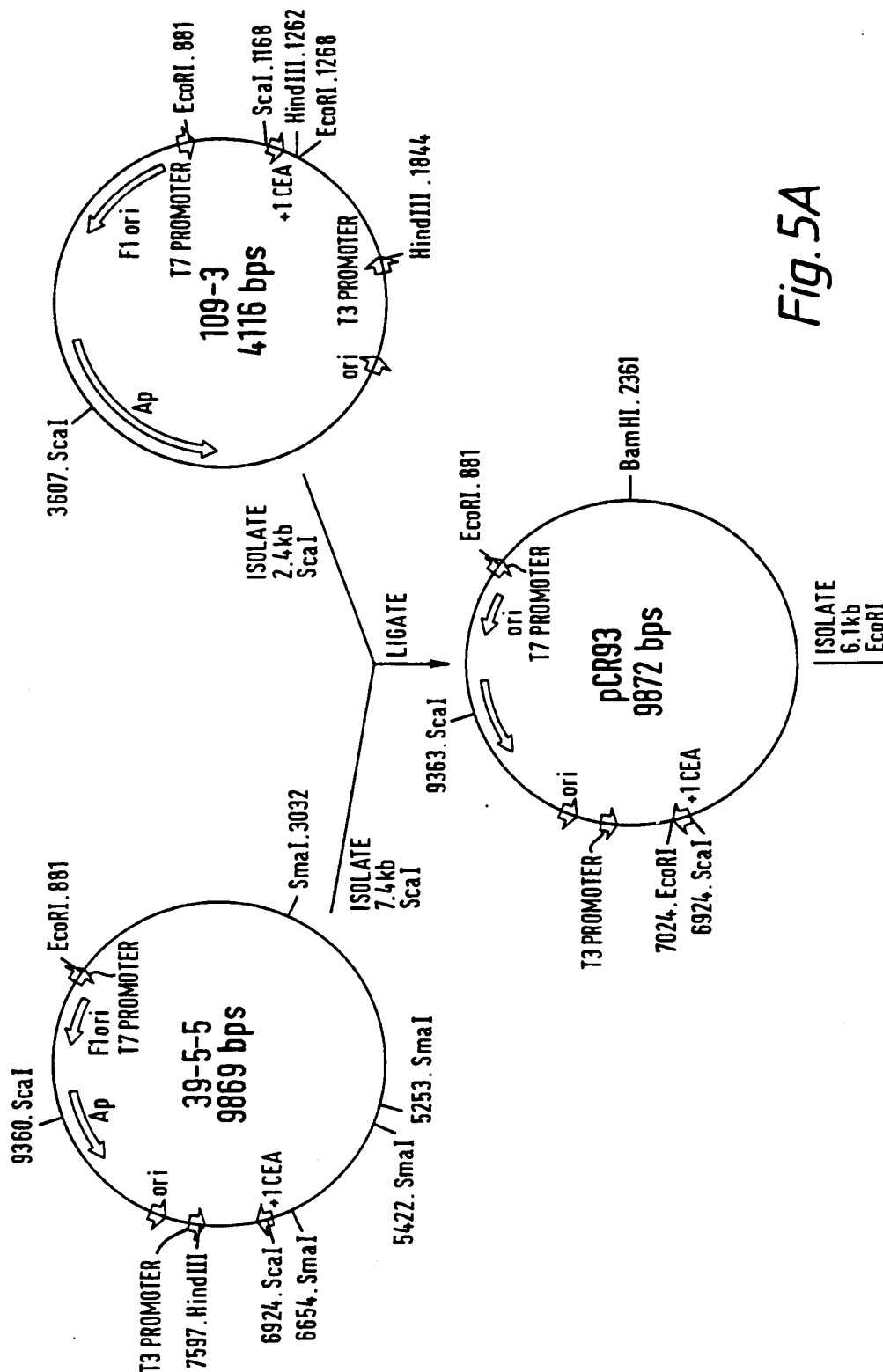


Fig. 5A

7/20

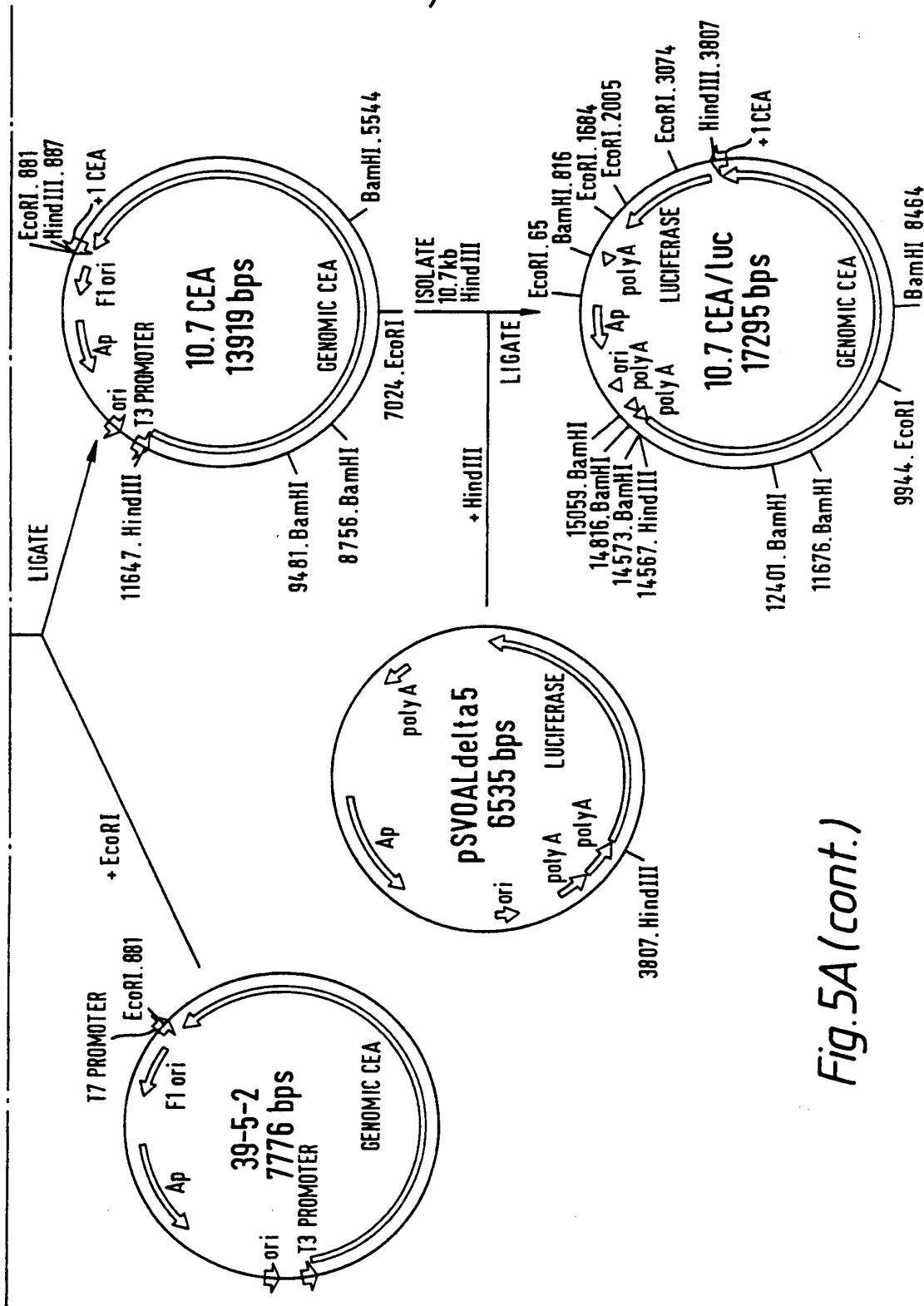


Fig. 5A (cont.)

8/20

<u>Plasmid</u>	<u>CEA Coordinates</u>
pCR113	(-299 to +69)
pCR105	(-1664 to +69)
pCR145	(-14462 to -10691)+(-299 to +69)
pCR148	(-89 to -40)+(-90 to +69)
pCR158	[3 X (-89 to -40)]+(-90 to +69)
pCR136	(-3919 to -6071) + (-299 to +69)
pCR137	(-6071 to -3919) + (-299 to +69)
pCR162	(-13579 to -10691)+(-89 to -40)+(-90 to +69)
pCR163	(-10691 to -13579)+(-89 to -40)+(-90 to +69)

Fig.5B

9/20

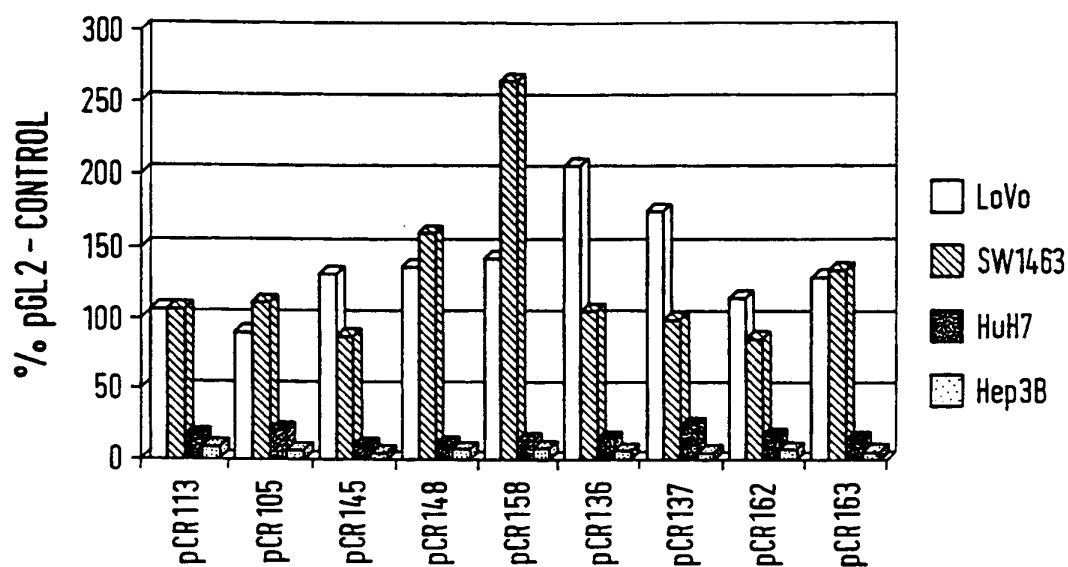


Fig.5C

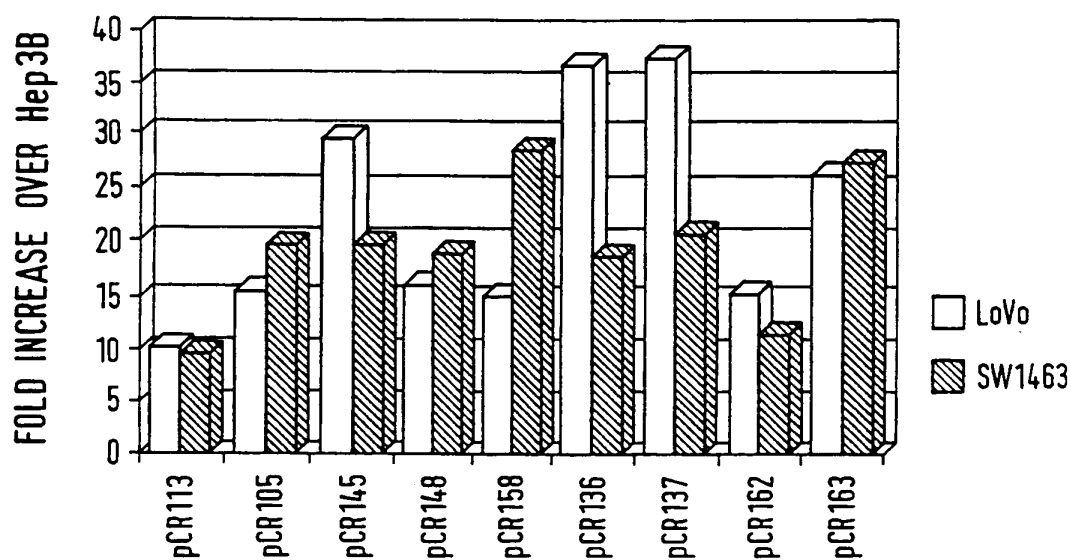


Fig.5D

10/20

-14463 AAGCTTTTTA GTGCTTTAGA CAGTGAGCTG GTCTGTCTAA CCCAAGTGAC CTGGGCTC
-14403 TACTCAGCCC CAGAAGTGAA GGGTGAAGCT GGGTGGAGCC AAACCAGGCA AGCCTACC
-14343 CAGGGCTCCC AGTGGCCTGA GAACCATTGG ACCCAGGACC CATTACTTCT AGGGTAAG
-14283 AGGTACAAAC ACCAGATCCA ACCATGGTCT GGGGGGACAG CTGTCAAATG CCTAAAAA
-14223 TACCTGGGAG AGGAGCAGGC AACTATCAC TGCCCCAGGT TCTCTGAACA GAAACAGA
-14163 GGCAACCCAA AGTCCAAATC CAGGTGAGCA GGTGCACCAA ATGCCCAGAG ATATGACG
-14103 GCAAGAAGTG AAGGAACCAC CCCTGCATCA AATGTTTTGC ATGGGAAGGA GAAGGGGG
-14043 GTCATGTTC CCAATCCAGG AGAATGCATT TGGGATCTGC CTTCTTCTCA CTCCTTGG
-13983 AGCAAGACTA AGCAACCAGG ACTCTGGATT TGGGGAAAGA CGTTTATTTG TGGAGGCC
-13923 TGATGACAAT CCCACGAGGG CCTAGGTGAA GAGGGCAGGA AGGCTCGAGA CACTGGGG
-13863 TGAGTGAAAA CCACACCCAT GATCTGCACC ACCCATGGAT GCTCCTTCAT TGCTCACC
-13803 TCTGTTGATA TCAGATGGCC CCATTTTCTG TACCTTCACA GAAGGACACA GGCTAGGG
-13743 TGTGCATGGC CTTTCATCCCC GGGGCCATGT GAGGACAGCA GGTGGGAAAG ATCATGGG
-13683 CTCCTGGGTC CTGCAGGGCC AGAACATTCA TCACCCATAC TGACCTCCTA GATGGGAA
-13623 GCTTCCCTGG GGCTGGGCCA ACGGGGCTG GGCAGGGGAG AAAGGACGTC AGGGGACA
-13563 GAGGAAGGGT CATCGAGACC CAGCCTGGAA GGTTCCTGTC TCTGACCATC CAGGATTT
-13503 TTCCCTGCAT CTACCTTTGG TCATTTTCCC TCAGCAATGA CCAGCTCTGC TTCCTGAT
-13443 CAGCCTCCCA CCCTGGACAC AGCACCCCAG TCCCTGGCCC GGCTGCATCC ACCCAATA
-13383 CTGATAACCC AGGACCCATT ACTTCTAGGG TAAGGAGGGT CCAGGAGACA GAAGCTGA
-13323 AAAGGTCTGA AGAAGTCACA TCTGTCCTGG CCAGAGGGGA AAAACCATCA GATGCTGA
-13263 CAGGAGAATG TTGACCCAGG AAAGGGACCG AGGACCCAAG AAAGGAGTCA GACCACCA
-13203 GTTTCCTGA GAGGAAGGAT CAAGGCCCCG AGGGAAAGCA GGGCTGGCTG CATGTGCA

Fig. 6 (1/11)

11/20

-13143 ACACTGGTGG GGCATATGTG TCTTAGATTC TCCCTGAATT CAGTGTCCCT GCCATGGC
-13083 GACTCTCTAC TCAGGCCTGG ACATGCTGAA ATAGGACAAT GGCCTTGTCC TCTCTCCC
-13023 CCATTTGGCA AGAGACATAA AGGACATTCC AGGACATGCC TTCCTGGGAG GTCCAGGT
-12963 TCTGTCTCAC ACCTCAGGGA CTGTAGTTAC TGCATCAGCC ATGGTAGGTG CTGATCTC
-12903 CCAGCCTGTC CAGGCCCTTC CACTCTCCAC TTTGTGACCA TGTCCAGGAC CACCCCTC
-12843 ATCCTGAGCC TGCAAATACC CCCTTGCTGG GTGGGTGGAT TCAGTAAACA GTGAGCTC
-12783 ATCCAGCCCC CAGAGCCACC TCTGTACCT TCCTGCTGGG CATCATCCCA CCTTCACA
-12723 CACTAAAGAG CATGGGGAGA CCTGGCTAGC TGGGTTTCTG CATCACAAAG AAAATAAT
-12663 CCCAGGTTCG GATTCCCAGG GCTCTGTATG TGGAGCTGAC AGACCTGAGG CCAGGAGA
-12603 GCAGAGGTCA GCCCTAGGGA GGGTGGGTCA TCCACCCAGG GGACAGGGGT GCACCAGC
-12543 TGCTACTGAA AGGGCCTCCC CAGGACAGCG CCATCAGCCC TGCCTGAGAG CTTTGCTA
-12483 CAGCAGTCAG AGGAGGCCAT GGCAGTGGCT GAGCTCCTGC TCCAGGCCCC AACAGACC
-12423 ACCAACAGCA CAATGCAGTC CTTCCCCAAC GTCACAGGTC ACCAAAGGGA AACTGAGG
-12363 CTACCTAACC TTAGAGCCAT CAGGGGAGAT AACAGCCCA TTTCCCAAAC AGGCCAGT
-12303 CAATCCCATG ACAATGACCT CTCTGCTCTC ATTCTTCCCA AAATAGGACG CTGATTCT
-12243 CCCACCATGG ATTTCTCCCT TGTCCCGGGA GCCTTTTCTG CCCCCTATGA TCTGGGCA
-12183 CCTGACACAC ACCTCCTCTC TGGTGACATA TCAGGGTCCC TCACTGTCAA GCAGTCCA
-12123 AAGGACAGAA CCTTGACAG CGCCCATCTC AGCTTCACCC TTCCTCCTTC ACAGGGTT
-12063 GGGCAAAGAA TAAATGGCAG AGGCCAGTGA GCCCAGAGAT GGTGACAGGC AGTGACCC
-12003 GGGCAGATGC CTGGAGCAGG AGCTGGCGGG GCCACAGGGA GAAGGTGATG CAGGAAGG
-11943 AAGCCAGAAA TGGGCAGGAA AGGAGGACAC AGGCTCTGTG GGGCTGCAGC CCAGGGTT
-11883 ACTATGAGTG TGAAGCCATC TCAGCAAGTA AGGCCAGGTC CCATGAACAA GAGTGGGA
-11823 ACGTGGCTTC CTGCTCTGTA TATGGGGTGG GGGATTCCAT GCCCCATAGA ACCAGATG

Fig. 6 (2/11)

12/20

-11763 CGGGGTTCAG ATGGAGAAGG AGCAGGACAG GGGATCCCCA GGATAGGAGG ACCCCAGT
-11703 CCCCACCCAG GCAGGTGACT GATGAATGGG CATGCAGGGT CCTCCTGGGC TGGGCTCT
-11643 CTTTGTCCCT CAGGATTCCT TGAAGGAACA TCCGGAAGCC GACCACATCT ACCTGGTG
-11583 TTCTGGGGAG TCCATGTAAA GCCAGGAGCT TGTGTTGCTA GGAGGGGTCA TGGCATGT
-11523 TGGGGGCACC AAAGAGAGAA ACCTGAGGGC AGGCAGGACC TGGTCTGAGG AGGCATGG
-11463 GCCCAGATGG GGAGATGGAT GTCAGGAAAG GCTGCCCCAT CAGGGAGGGT GATAGCAA
-11403 GGGGTCTGT GGGAGTGGGC ACGTGGGATT CCCTGGGCTC TGCCAAGTTC CCTCCCAT
-11343 TCACAACCTG GGGACACTGC CCATGAAGGG GCGCCTTTGC CCAGCCAGAT GCTGCTGG
-11283 CTGCCCATCC ACTACCCTCT CTGCTCCAGC CACTCTGGGT CTTTCTCCAG ATGCCCTG
-11223 CAGCCCTGGC CTGGGCCTGT CCCCTGAGAG GTGTTGGGAG AAGCTGAGTC TCTGGGGA
-11163 CTCTCATCAG AGTCTGAAAG GCACATCAGG AAACATCCCT GGTCTCCAGG ACTAGGCA
-11103 GAGGAAAGGG CCCCAGCTCC TCCCTTTGCC ACTGAGAGGG TCGACCCTGG GTGGCCAC
-11043 TGACTTCTGC GTCTGTCCCA GTCACCCTGA AACCACAACA AAACCCAGC CCCAGACC
-10983 GCAGGTACAA TACATGTGGG GACAGTCTGT ACCCAGGGGA AGCCAGTTCT CTCTTCCT
-10923 GAGACCGGGC CTCAGGGCTG TGCCCGGGC AGGCGGGGC AGCACGTGCC TGTCCCTG
-10863 AACTCGGGAC CTTAAGGGTC TCTGCTCTGT GAGGCACAGC AAGGATCCTT CTGTCCAG
-10803 ATGAAAGCAG CTCCTGCCCC TCCTCTGACC TCTTCCTCCT TCCCAAATCT CAACCAAC
-10743 ATAGGTGTTT CAAATCTCAT CATCAAATCT TCATCCATCC ACATGAGAAA GCTTAAAA
-10683 CAATGGATTG ACAACATCAA GAGTTGGAAC AAGTGGACAT GGAGATGTTA CTTGTGGA
-10623 TTTAGATGTG TTCAGCTATC GGGCAGGAGA ATCTGTGTCA AATTCCAGCA TGGTTCAG
-10563 GAATCAAAAA GTGTCACAGT CCAAATGTGC AACAGTGCAG GGGATAAAAC TGTGGTGC
-10503 TCAAACCTGAG GGATATTTTG GAACATGAGA AAGGAAGGGA TTGCTGCTGC ACAGAACA
-10443 GATGATCTCA CACATAGAGT TGAAAGAAAG GAGTCAATCG CAGAATAGAA AATGATCA

Fig. 6 (3/11)

13/20

-10383 AATTCCACCT CTATAAAGTT TCCAAGAGGA AAACCCAATT CTGCTGCTAG AGATCAGA
-10323 GGAGGTGACC TGTGCCTTGC AATGGCTGTG AGGGTCACGG GAGTGTCACT TAGTGCAG
-10263 AATGTGCCGT ATCTTAATCT GGGCAGGGCT TTCATGAGCA CATAGGAATG CAGACATT
-10203 TGCTGTGTTT ATTTTACTTC ACCGGAAG AAGAATAAAA TCAGCCGGGC GCGGTGGC
-10143 ACGCCTGTAA TCCCAGCACT TTAGAAGGCT GAGGTGGGCA GATTACTTGA GGTGAGGA
-10083 TCAAGACCAC CCTGGCCAAT ATGGTGAAAC CCCGGCTCTA CTAAAAATAC AAAAATTA
-10023 TGGGCATGGT GGTGCGCGCC TGTAATCCCA GCTACTCGGG AGGCTGAGGC TGGACAAT
-9963 CTTGGACCCA GGAAGCAGAG GTTGCACTGA GCCAAGATTG TGCCACTGCA CTCCAGCT
-9903 GGCAACAGAG CCAGACTCTG TAAAAA AAAAAGAA AGAAAGAA
-9843 AGAAAGAAA GTATAAATC TCTTTGGGT AACAAAAA GATCCACAA ACAACAC
-9783 GCTCTTATCA AACTTACACA ACTCTGCCAG AGAACAGGAA ACACAAATAC TCATTAAC
-9723 ACTTTTGTGG CAATAAACC TTCATGTCAA AAGGAGACCA GGACACAATG AGGAAGTA
-9663 ACTGCAGGCC CTACTTGGGT GCAGAGAGGG AAAATCCACA AATAAACAT TACCAGAA
-9603 AGCTAAGATT TACTGCATTG AGTTCATTCC CCAGGTATGC AAGGTGATT TAACACCT
-9543 AAATCAATCA TTGCCTTTAC TACATAGACA GATTAGCTAG AAAAAATTA CAACTAGC
-9483 AACAGAAGCA ATTTGGCCTT CCTAAAATC CACATCATAT CATCATGATG GAGACAGT
-9423 AGACGCCAAT GACAATAAAA AGAGGGACCT CCGTCACCCG GTAAACATGT CCACACAG
-9363 CCAGCAAGCA CCCGTCTTCC CAGTGAATCA CTGTAACCTC CCCTTTAATC AGCCCCAG
-9303 AAGGCTGCCT GCGATGGCCA CACAGGCTCC AACCCGTGGG CCTCAACCTC CCGCAGAG
-9243 TCTCCTTTGG CCACCCCATG GGGAGAGCAT GAGGACAGGG CAGAGCCCTC TGATGCCC
-9183 ACATGGCAGG AGCTGACGCC AGAGCCATGG GGGCTGGAGA GCAGAGCTGC TGGGGTCA
-9123 GCTTCCTGAG GACACCCAGG CCTAAGGGAA GGCAGCTCCC TGGATGGGG CAACCAGG
-9063 CCGGGCTCCA ACCTCAGAGC CCGCATGGGA GGAGCCAGCA CTCTAGGCCT TTCCTAGG

Fig. 6 (4/11)

14/20

-9003 GACTCTGAGG GGACCCTGAC ACGACAGGAT CGCTGAATGC ACCCGAGATG AAGGGGCC
-8943 CACGGGACCC TGCTCTCGTG GCAGATCAGG AGAGAGTGGG ACACCATGCC AGGCCCCC
-8883 GGCATGGCTG CGACTGACCC AGGCCACTCC CCTGCATGCA TCAGCCTCGG TAAGTCAC
-8823 GACCAAGCCC AGGACCAATG TGGAAGGAAG GAAACAGCAT CCCCTTTAGT GATGGAAC
-8763 AAGGTCAGTG CAAAGAGAGG CCATGAGCAG TTAGGAAGGG TGGTCCAACC TACAGCAC
-8703 ACCATCATCT ATCATAAGTA GAAGCCCTGC TCCATGACCC CTGCATTAA ATAAACGT
-8643 GTTAAATGAG TCAAATTCCC TCACCATGAG AGCTCACCTG TGTGTAGGCC CATCACAC
-8583 ACAAACACAC ACACACACAC ACACACACAC ACACACACAC ACAGGGAAAG TGCAGGAT
-8523 TGGACAGCAC CAGGCAGGCT TCACAGGCAG AGCAAACAGC GTGAATGACC CATGCAGT
-8463 CCTGGGCCCC ATCAGCTCAG AGACCCTGTG AGGGCTGAGA TGGGGCTAGG CAGGGGAG
-8403 ACTTAGAGAG GGTGGGGCCT CCAGGGAGGG GGCTGCAGGG AGCTGGGTAC TGCCCTCC
-8343 GGAGGGGGCT GCAGGGAGCT GGGTACTGCC CTCCAGGGAG GGGGCTGCAG GGAGCTGG
-8283 ACTGCCCTCC AGGGAGGGGG CTGCAGGGAG CTGGGTACTG CCCTCCAGGG AGGGGGCT
-8223 AGGGAGCTGG GTACTGCCCT CCAGGGAGGC AGGAGCACTG TTCCCAACAG AGAGCACA
-8163 TTCCTGCAGC AGCTGCACAG ACACAGGAGC CCCCATGACT GCCCTGGGCC AGGGTGTG
-8103 TTCCAAATTT CGTGCCCCAT TGGGTGGGAC GGAGGTTGAC CGTGACATCC AAGGGGCA
-8043 TGTGATTCCA AACTTAAACT ACTGTGCCTA CAAAATAGGA AATAACCCTA CTTTTTCT
-7983 TATCTCAAAT TCCCTAAGCA CAAGCTAGCA CCCTTTAAAT CAGGAAGTTC AGTCACTC
-7923 GGGGTCCTCC CATGCCCCCA GTCTGACTTG CAGGTGCACA GGGTGGCTGA CATCTGTC
-7863 TGCTCCTCCT CTTGGCTCAA CTGCCGCCCC TCCTGGGGGT GACTGATGGT CAGGACAA
-7803 GATCCTAGAG CTGGCCCCAT GATTGACAGG AAGGCAGGAC TTGGCCTCCA TTCTGAAG
-7743 TAGGGGTGTC AAGAGAGCTG GGCATCCAC AGAGCTGCAC AAGATGACGC GGACAGAG
-7683 TGACACAGGG CTCAGGGCTT CAGACGGGTC GGGAGGCTCA GCTGAGAGTT CAGGGACA

Fig. 6 (5/11)

15/20

-7623 CCTGAGGAGC CTCAGTGGGA AAAGAAGCAC TGAAGTGGGA AGTTCTGGAA TGTCTCTGG
-7563 AAGCCTGAGT GCTCTAAGGA AATGCTCCCA CCCCAGATGTA GCCTGCAGCA CTGGACGG
-7503 TGTGTACCTC CCCGCTGCCC ATCCTCTCAC AGCCCCCGCC TCTAGGGACA CAACTCCT
-7443 CCTAACATGC ATCTTTCCTG TCTCATTCCA CACAAAAGGG CCTCTGGGGT CCCTGTTC
-7383 CATTGCAAGG AGTGGAGGTC ACGTTCCAC AGACCACCCA GCAACAGGGT CCTATGGA
-7323 TGCGGTCAGG AGGATCACAC GTCCCCCAT GCCCAGGGGA CTGACTCTGG GGGTGATG
-7263 TTGGCCTGGA GGCCACTGGT CCCCTCTGTC CCTGAGGGGA ATCTGCACCC TGGAGGCT
-7203 CACATCCCTC CTGATTCTTT CAGCTGAGGG CCCTTCTTGA AATCCCAGGG AGGACTCA
-7143 CCCCCTGGG AAAGGCCAG TGTGGACGGT TCCACAGCAG CCCAGCTAAG GCCCTTGG
-7083 ACAGATCCTG AGTGAGAGAA CCTTTAGGGA CACAGGTGCA CGGCCATGTC CCCAGTGC
-7023 ACACAGAGCA GGGGCATCTG GACCCTGAGT GTGTAGCTCC CGCGACTGAA CCCAGCCC
-6963 CCCCAATGAC GTGACCCCTG GGGTGGCTCC AGGTCTCCAG TCCATGCCAC CAAAATCT
-6903 AGATTGAGGG TCCTCCCTTG AGTCCCTGAT GCCTGTCCAG GAGCTGCCCC CTGAGCAA
-6843 CTAGAGTGTA GAGGGCTGGG ATTGTGGCAG TAAAAGCAGC CACATTTGTC TCAGGAAG
-6783 AAGGGAGGAC ATGAGCTCCA GGAAGGGCGA TGGCGTCCTC TAGTGGGCGC CTCCTGTT
-6723 TGAGCAAAAA GGGGCCAGGA GAGTTGAGAG ATCAGGGCTG GCCTTGGA CT AAGGCTCA
-6663 TGGAGAGGAC TGAGGTGCAA AGAGGGGGCT GAAGTAGGGG AGTGGTCGGG AGAGATGG
-6603 GGAGCAGGTA AGGGGAAGCC CCAGGGAGGC CGGGGGAGGG TACAGCAGAG CTCTCCAC
-6543 CTCAGCATTG ACATTTGGGG TGGTCGTGCT AGTGGGGTTC TGTAAGTTGT AGGGTGTT
-6483 GCACCATCTG GGGACTCTAC CCACTAAATG CCAGCAGGAC TCCCTCCCCA AGCTCTAA
-6423 ACCAACAATG TCTCCAGACT TTCCAAATGT CCCCTGGAGA GCAAAATTGC TTCTGGCA
-6363 ATCACTGATC TACGTCAGTC TCTAAAAGTG ACTCATCAGC GAAATCCTTC ACCTCTTG
-6303 AGAAGAATCA CAAGTGTGAG AGGGGTAGAA ACTGCAGACT TCAAAATCTT TCCAAAAG

Fig. 6 (6/11)

16/20

-6243 TTTTACTTAA TCAGCAGTTT GATGTCCCAG GAGAAGATAC ATTTAGAGTG TTTAGAGT
-6183 ATGCCACATG GCTGCCTGTA CCTCACAGCA GGAGCAGAGT GGGTTTTCCA AGGGCCTG
-6123 ACCACAAC TG GAATGACACT CACTGGGT TA CATTACAAAG TGGAATGTGG GGAATTCT
-6063 AGACTTTGGG AAGGGAAATG TATGACGTGA GCCCACAGCC TAAGGCAGTG GACAGTCC
-6003 TTTGAGGCTC TCACCATCTA GGAGACATCT CAGCCATGAA CATAGCCACA TCTGTCAT
-5943 GAAAACATGT TTTATTAAGA GGAAAAATCT AGGCTAGAAG TGCTTTATGC TCTTTTTT
-5883 CTTTATGTTC AAATTCATAT ACTTTTAGAT CATTCCTTAA AGAAGAATCT ATCCCCCT
-5823 GTAAATGTTA TCACTGACTG GATAGTGTG GTGTCTCACT CCCAACCCT GTGTGGTG
-5763 AGTGCCCTGC TTCCCCAGCC CTGGGCCCTC TCTGATTCCT GAGAGCTTTG GGTGCTCC
-5703 CATTAGGAGG AAGAGAGGAA GGGTGTMTT AATATTCTCA CCATTCACCC ATCCACCT
-5643 TAGACACTGG GAAGAATCAG TTGCCCACTC TTGGATTGA TCCTCGAATT AATGACCT
-5583 ATTTCTGTCC CTTGTCCATT TCAACAATGT GACAGGCCTA AGAGGTGCCT TCTCCATG
-5523 ATTTTTGAGG AGAAGGTTCT CAAGATAAGT TTTCTCACAC CTCTTTGAAT TACCTCCA
-5463 TGTGTCCCCA TCACCATAC CAGCAGCATT TGGACCCTT TTCTGTTAGT CAGATGCT
-5403 CCACCTCTTG AGGGTGTATA CTGTATGCTC TCTACACAGG AATATGCAGA GGAAATAG
-5343 AAAGGGAAAT CGCATTACTA TTCAGAGAGA AGAAGACCTT TATGTGAATG AATGAGAG
-5283 TAAAATCCTA AGAGAGCCCA TATAAATTA TTACCAGTGC TAAAACTACA AAAGTTAC
-5223 TAACAGTAAA CTAGAATAAT AAAACATGCA TCACAGTTGC TGGTAAAGCT AAATCAGA
-5163 TTTTTTTCTT AGAAAAAGCA TTCCATGTGT GTTGCAAGTGA TGACAGGAGT GCCCTTCA
-5103 CAATATGCTG CCTGTAATTT TTGTTCCCTG GCAGAATGTA TTGTCTTTTC TCCCTTTA
-5043 TCTTAAATGC AAAACTAAAG GCAGCTCCTG GGCCCCCTCC CCAAAGTCAG CTGCCTGC
-4983 CCAGCCCCAC GAAGAGCAGA GGCCTGAGCT TCCCTGGTCA AAATAGGGGG CTAGGGAG
-4923 TAACCTTGCT CGATAAAGCT GTGTTCCCAG AATGTCGCTC CTGTTCCCAG GGGCACCA

Fig. 6 (7/11)

17/20

-4863 CTGGAGGGTG GTGAGCCTCA CTGGTGGCCT GATGCTTACC TTGTGCCCTC ACACCACT
-4803 TCACTGGAAC CTTGAACACT TGGCTGTCGC CCGGATCTGC AGATGTCAAG AACTTCTG
-4743 AGTCAAATTA CTGCCCACCTT CTCCAGGGCA GATACCTGTG AACATCCAAA ACCATGCC
-4683 AGAACCCTGC CTGGGGTCTA CAACACATAT GGA CTGTGAG CACCAAGTCC AGCCCTGA
-4623 CTGTGACCAC CTGCCAAGAT GCCCCTAACT GGGATCCACC AATCACTGCA CATGGCAG
-4563 AGCGAGGCTT GGAGGTGCTT CGCCACAAGG CAGCCCCAAT TTGCTGGGAG TTTCTTGG
-4503 CCTGGTAGTG GTGAGGAGCC TTGGGACCCT CAGGATTACT CCCCTTAAGC ATAGTGGG
-4443 CCCTTCTGCA TCCCCAGCAG GTGCCCCGCT CTTGAGAGCC TCTCTCTCTG AGGTTTAC
-4383 AGACCCCTGC ACCAATGAGA CCATGCTGAA GCCTCAGAGA GAGAGATGGA GCTTTGAC
-4323 GGAGCCGCTC TTCCTTGAGG GCCAGGGCAG GGAAAGCAGG AGGCAGCACC AGGAGTGG
-4263 ACACCAGTGT CTAAGCCCCT GATGAGAACA GGGTGGTCTC TCCCATATGC CCATACCA
-4203 CCTGTGAACA GAATCCTCCT TCTGCAGTGA CAATGTCTGA GAGGACGACA TGTTTTCC
-4143 CCTAACGTGC AGCCATGCCC ATCTACCCAC TGCCTACTGC AGGACAGCAC CAACCCAG
-4083 GCTGGGAAGC TGGGAGAAGA CATGGAFTAC CCATGGCTTC TCACCTTCCT CCAGTCCA
-4023 GGGCACCATT TATGCCTAGG ACACCCACCT GCCGGCCCCA GGCTCTTAAG AGTTAGGT
-3963 CCTAGGTGCC TCTGGGAGGC CGAGGCAGGA GAATTGCTTG AACC CGGGAG GCAGAGGT
-3903 CAGTGAGCCG AGATCACACC ACTGCACTCC AGCCTGGGTG ACAGAATGAG ACTCTGTC
-3843 AAAAAAAAAAG AGAAAGATAG CATCAGTGGC TACCAAGGGC TAGGGGCAGG GGAAGGTG
-3783 GAGTTAATGA TTAATAGTAT GAAGTTTCTA TGTGAGATGA TGAAAATGTT CTGGAAAA
-3723 AAATATAGTG GTGAGGATGT AGAATATTGT GAATATAATT AACGGCATT AATTGTAC
-3663 TTAACATGAT TAATGTGGCA TATTTTATCT TATGTATTG ACTACATCCA AGAAACAC
-3603 GGAGAGGGAA AGCCCACCAT GTAAAATACA CCCACCCTAA TCAGATAGTC CTCATTGT
-3543 CCAGGTACAG GCCCCTCATG ACCTGCACAG GAATAACTAA GGATTTAAGG ACATGAGG

Fig. 6 (8/11)

18/20

-3483 TCCCAGCCAA CTGCAGGTGC ACAACATAAA TGTATCTGCA AACAGACTGA GAGTAAAG
-3423 GGGGGCACAA ACCTCAGCAC TGCCAGGACA CACACCCTTC TCGTGGATTC TGACTTTA
-3363 TGACCCGGCC CACTGTCCAG ATCTTGTTGT GGGATTGGGA CAAGGGAGGT CATAAAGC
-3303 GTCCCCAGGG CACTCTGTGT GAGCACACGA GACCTCCCCA CCCCCCACC GTTAGGTC
-3243 CACACATAGA TCTGACCATT AGGCATTGTG AGGAGGACTC TAGCGCGGGC TCAGGGAT
-3183 CACCAGAGAA TCAGGTACAG AGAGGAAGAC GGGGCTCGAG GAGCTGATGG ATGACACA
-3123 GCAGGGTTCC TGCAGTCCAC AGGTCCAGCT CACCCTGGTG TAGGTGCCCC ATCCCCCT
-3063 TCCAGGCATC CCTGACACAG CTCCCTCCCG GAGCCTCCTC CCAGGTGACA CATCAGGG
-3003 CCTCACTCAA GCTGTCCAGA GAGGGCAGCA CCTTGGACAG CGCCCACCCC ACTTCACT
-2943 TCCTCCCTCA CAGGGCTCAG GGCTCAGGGC TCAAGTCTCA GAACAAATGG CAGAGGCC
-2883 TGAGCCCAGA GATGGTGACA GGGCAATGAT CCAGGGGCAG CTGCCTGAAA CGGGAGCA
-2823 TGAAGCCACA GATGGGAGAA GATGGTTCAG GAAGAAAAT CCAGGAATGG GCAGGAGA
-2763 AGAGGAGGAC ACAGGCTCTG TGGGGCTGCA GCCCAGGATG GGAATAAGTG TGAAGACA
-2703 TCAGCAGGTG AGGCCAGGTC CCATGAACAG AGAAGCAGCT CCCACCTCCC CTGATGCA
-2643 GACACACAGA GTGTGTGGTG CTGTGCCCCC AGAGTCGGGC TCTCCTGTTT TGGTCCCC
-2583 GGAGTGAGAA GTGAGGTTGA CTTGTCCCTG CTCCTCTCTG CTACCCCAAC ATTCACCT
-2523 TCCTCATGCC CCTCTCTCTC AAATATGATT TGGATCTATG TCCCCGCCCA AATCTCAT
-2463 CAAATTGTAA ACCCCAATGT TGGAGGTGGG GCCTTGTGAG AAGTGATTGG ATAATGCG
-2403 TGGATTTTCT GCTTTGATGC TGTTTCTGTG ATAGAGATCT CACATGATCT GGTTGTTT
-2343 AAGTGTGTAG CACCTCTCCC CTCTCTCTCT CTCTCTCTTA CTCATGCTCT GCCATGTA
-2283 ACGTTCCTGT TTCCCTTCA CCGTCCAGAA TGATTGTAAG TTTTCTGAGG CCTCCCCA
-2223 AGCAGAAGCC ACTATGCTTC CTGTACAACT GCAGAATGAT GAGCGAATTA AACCTCTT
-2163 CTTTATAAAT TACCCAGTCT CAGGTATTTT TTTATAGCAA TGCGAGGACA GACTAATA

Fig. 6 (9/11)

19/20

-2103 ATCTTCTACT CCCAGATCCC CGCACACGCT TAGCCCCAGA CATCACTGCC CCTGGGAG
-2043 TGCACAGCGC AGCCTCCTGC CGACAAAAGC AAAGTCACAA AAGGTGACAA AAATCTGC
-1983 TTGGGGACAT CTGATTGTGA AAGAGGGAGG ACAGTACACT TGTAGCCACA GAGACTGG
-1923 CTCACCGAGC TGAAACCTGG TAGCACTTTG GCATAACATG TGCATGACCC GTGTTCAA
-1863 TCTAGAGATC AGTGTGTAGT AAAACAGCCT GGTCTGGGGC CGCTGCTGTC CCCACTTC
-1803 TCCTGTCCAC CAGAGGGCGG CAGAGTTCCT CCCACCCTGG AGCCTCCCCA GGGGCTGC
-1743 ACCTCCCTCA GCCGGGCCCA CAGCCCAGCA GGGTCCACCC TCACCCGGGT CACCTCGG
-1683 CACGTCCTCC TCGCCCTCCG AGCTCCTCAC ACGGACTCTG TCAGCTCCTC CCTGCAGC
-1623 ATCGGCCGCC CACCTGAGGC TTGTCGGCCG CCCACTTGAG GCCTGTCCGC TGCCCTCT
-1563 AGGCAGCTCC TGTCCCCTAC ACCCCCTCCT TCCCCGGGCT CAGCTGAAAG GGCCTCTC
-1503 AGGGCAGCTC CCTGTGATCT CCAGGACAGC TCAGTCTCTC ACAGGCTCCG ACGCCCCC
-1443 TGCTGTCACC TCACAGCCCT GTCATTACCA TTAATCCTC AGTCCCATGA AGTTCACT
-1383 GCGCCTGTCT CCCGGTTACA GGAAACTCT GTGACAGGGA CCACGTCTGT CTGCTCT
-1323 GTGGAATCCC AGGGCCCAGC CCAGTGCCTG ACACGGAACA GATGCTCCAT AAATACTG
-1263 TAAATGTGTG GGAGATCTCT AAAAAGAAGC ATATCACCTC CGTGTGGCCC CCAGCAGT
-1203 GAGTCTGTTC CATGTGGACA CAGGGGCACT GGCACCAGCA TGGGAGGAGG CCAGCAAG
-1143 CCCGCGGCTG CCCCAGGAAT GAGGCCTCAA CCCCAGAGC TTCAGAAGGG AGGACAGA
-1083 CCTGCAGGGA ATAGATCCTC CGGCCTGACC CTGCAGCCTA ATCCAGAGTT CAGGGTCA
-1023 TCACACCACG TCGACCCTGG TCAGCATCCC TAGGGCAGTT CCAGACAAGG CCGGAGGT
-963 CCTCTTGCCC TCCAGGGGGT GACATTGCAC ACAGACATCA CTCAGGAAAC GGATTCCC
-903 GGACAGGAAC CTGGCTTTGC TAAGGAAGTG GAGGTGGAGC CTGGTTTCCA TCCCTTGC
-843 CAACAGACCC TTCTGATCTC TCCCACATAC CTGCTCTGTT CCTTCTGGG TCCTATGA
-783 ACCCTGTTCT GCCAGGGGTC CCTGTGCAAC TCCAGACTCC CTCCTGGTAC CACCATGG

Fig. 6 (10/11)

20/20

-723 AAGGTGGGGT GATCACAGGA CAGTCAGCCT CGCAGAGACA GAGACCACCC AGGACTGT
-663 GGGAGAACAT GGACAGGCCC TGAGCCGCAG CTCAGCCAAC AGACACGGAG AGGGAGGG
-603 CCCCTGGAGC CTTCCCCAAG GACAGCAGAG CCCAGAGTCA CCCACCTCCC TCCACCAC
-543 TCCTCTCTTT CCAGGACACA CAAGACACCT CCCCCTCCAC ATGCAGGATC TGGGGACT
-483 TGAGACCTCT GGGCCTGGGT CTCCATCCCT GGGTCAGTGG CGGGGTGGT GGTACTGG
-423 ACAGAGGGCT GGTCCCTCCC CAGCCACCAC CCAGTGAGCC TTTTCTAGC CCCCAGAG
-363 ACCTCTGTCA CCTTCCTGTT GGGCATCATC CCACCTTCCC AGAGCCCTGG AGAGCATG
-303 GAGACCCGGG ACCCTGCTGG GTTCTCTGT CACAAAGGAA AATAATCCCC CTGGTGTG
-243 AGACCCAAGG ACAGAACACA GCAGAGGTCA GCACTGGGGA AGACAGGTTG TCCTCCCA
-183 GGATGGGGGT CCATCCACCT TGCCGAAAAG ATTTGTCTGA GGAAGTAAA ATAGAAGG
-123 AAAAAGAGGA GGGACAAAAG AGGCAGAAAT GAGAGGGGAG GGGACAGAGG ACACCTGA
-63 AAAGACCACA CCCATGACCC ACGTGATGCT GAGAAGTACT CCTGCCCTAG GAAGAGAC
-3 AGGGCAGAGG GAGGAAGGAC AGCAGACCAG ACAGTCACAG CAGCCTTGAC AAAACGTT
57 TGGAAGTCAA GCTCTTCTCC ACAGAGGAGG ACAGAGCAGA CAGCAGAGAC CATGAGT
117 CCCTCGGCCC CTCCCCACAG ATGGTGCATC CCCTGGCAGA GGCTCCTGCT CACAGGTG
177 GGGAGGACAA CCTGGGAGAG GGTGGGAGGA GGGAGCTGGG GTCTCCTGGG TAGGACAG
237 CTGTGAGACG GACAGAGGGC TCCTGTTGGA GCCTGAATAG GGAAGAGGAC ATCAGAGA
297 GACAGGAGTC ACACCAGAAA AATCAAATTG AACTGGAATT GGAAAGGGGC AGGAAAAC
357 CAAGAGTTCT ATTTTCTAG TTAATTGTCA CTGGCCACTA CGTTTTTAAA AATCATAA
417 ACTGCATCAG ATGACACTTT AAATAAAAAC ATAACCAGGG CATGAAACAC TGTCTCA
477 CGCTACCGC GGACATTGGA AAATAAGCCC CAGGCTGTGG AGGGCCCTGG GAACCTC
537 GAACTCATCC ACAGGAATCT GCAGCCTGTC CCAGGCACTG GGGTGCAACC AAGATC

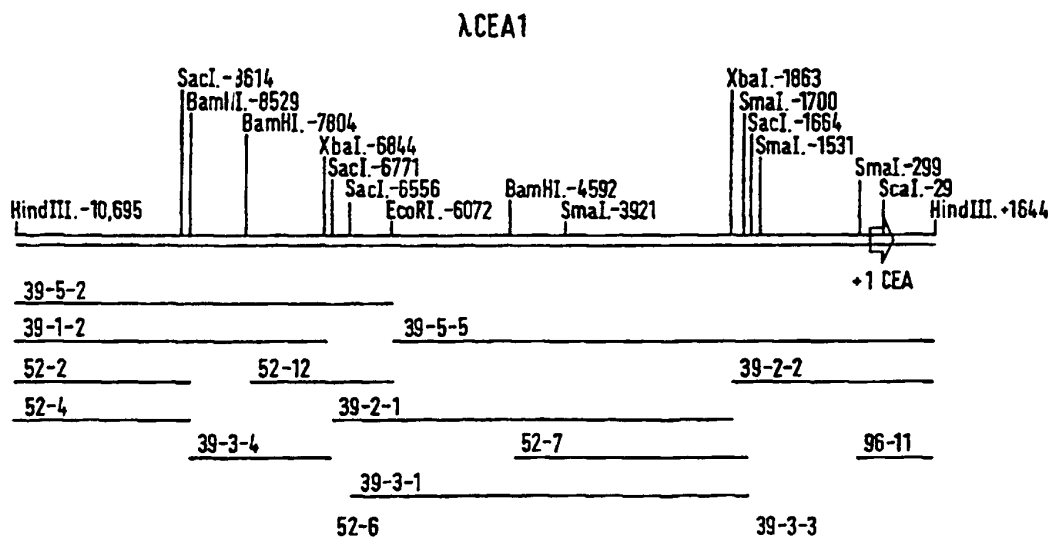
Fig.6 (11/11)

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/85, 15/86, 7/01, A61K 48/00, C12N 5/10		A3	(11) International Publication Number: WO 95/14100 (43) International Publication Date: 26 May 1995 (26.05.95)
(21) International Application Number: PCT/GB94/02546 (22) International Filing Date: 18 November 1994 (18.11.94) (30) Priority Data: 154,712 19 November 1993 (19.11.93) US (71) Applicant (for all designated States except US): THE WELL- COME FOUNDATION LIMITED [GB/GB]; Unicom House, 160 Euston Road, London NW1 2BP (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): RICHARDS, Cynthia, Ann [US/US]; 5616 Welkin Court, Durham, NC 27713 (US). HUBER, Brian [US/US]; 53 Westridge Drive, Durham, NC 27713 (US). (74) Agent: STOTT, Michael, John; The Wellcome Foundation Limited, Langley Court, Beckenham, Kent BR3 3BS (GB).			(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 15 June 1995 (15.06.95)

(54) Title: TRANSCRIPTIONAL REGULATORY SEQUENCE OF CARCINOEMBRYONIC ANTIGEN FOR EXPRESSION TARGET-
ING

(57) Abstract

The invention relates to the transcriptional regulatory sequence (TRS) of carcinoembryonic antigen (CEA) and molecular chimaera comprising the CEA TRS and DNA encoding a heterologous enzyme. CEA TRS is capable of targeting expression of the heterologous enzyme to CEA⁺ cells and the heterologous enzyme is preferably an enzyme capable of catalysing the production of an agent cytotoxic or cytostatic to CEA⁺ cells. For example the enzyme may be cytosine deaminase which is capable of catalysing formation of the cytotoxic compound 5-fluorouracil from the non toxic compound 5-fluorocytosine.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

INTERNATIONAL SEARCH REPORT

Inventor Application No

PCT/GB 94/02546

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/85 A61K48/00 C12N15/86 C12N5/10 C12N7/01

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DNA SEQUENCE, vol.4, no.3, July 1993, SWITZERLAND pages 185 - 196 RICHARDS, C. A. ET AL. 'The transcriptional control region of the human carcinoembryonic antigen gene: DNA sequence and homology studies'	1
Y	see the whole document --- -/--	2-19

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

* & * document member of the same patent family

Date of the actual completion of the international search

12 April 1995

Date of mailing of the international search report

11-05- 1995

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+ 31-70) 340-3016

Authorized officer

Chambonnet, F

INTERNATIONAL SEARCH REPORT

Application No
PCT/GB 94/02546

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MOLECULAR AND CELLULAR BIOLOGY, vol.10, no.6, June 1990 pages 2738 - 2748 SCHREWE, H. ET AL. 'Cloning of the complete gene for carcinoembryonic antigen : analysis of its promoter indicates a region conveying cell-type specific expression' cited in the application	1
Y	see the whole document see page 2739, column 1, line 1 - line 6; figure 4	1-19
P,X	--- ANNALS OF THE NEW YORK ACADEMY OF SCIENCES, vol.716, 1994 pages 104 - 114 HUBER, B.E. ET AL. 'Virus-Directed Enzyme/Prodrug Therapy (VDEPT) : Selectively engineering drug sensitivity into tumors' see the whole document	2-16
Y,P O,X	& Gene therapy for neoplastic diseases Washington, USA june 26-29th 1993	2-19 3
X	--- TRENDS IN BIOTECHNOLOGY., vol.11, no.5, May 1993, CAMBRIDGE GB pages 197 - 201 SIKORA, K. 'Gene therapy for cancer' see page 198, column 2, line 6 - page 199, column 1, line 34; figure 2; table 1	2-5, 14-16
Y	--- see page 198, column 2, line 6 - page 199, column 1, line 34; figure 2; table 1	2-19
P,Y	--- GENE THERAPY, vol.1, no.3, May 1994 pages 170 - 175 HARRIS, J. D. ET AL. 'Gene therapy for cancer using tumour-specific prodrug activation' see the whole document	2-19
Y	--- CANCER RESEARCH, vol.53, no.19, 1 October 1993 pages 4619 - 4626 HUBER, B.E. ET AL. 'In vivo antitumor activity of 5-fluorocytosine on human colorectal carcinoma cells genetically modified to express cytosine deaminase' see the whole document	2-19
	--- -/--	

INTERNATIONAL SEARCH REPORT

Application No
PCT/GB 94/02546

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	CANCER RESEARCH, vol.54, no.20, 15 October 1994 pages 5258 - 5261 TADASHI OSAKI ET AL. 'Gene therapy for carcinoembryonic antigen-producing human lung cancer cells by cell type-specific expression of Herpes Simplex Virus Thymidine Kinase gene'	2,3,5
Y	see page 5258, column 2, line 16 - line 34 see page 5261, column 1, line 6 - column 2, line 8	1-19
A	--- EP,A,0 415 731 (WELLCOME FOUNDATION LTD.) 6 March 1991 cited in the application see the whole document	1
A	--- WO,A,92 15693 (THE WELLCOME FOUNDATION) 17 September 1992 see claims 1-3 -----	1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CIB94/02546

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos. because they relate to subject matter not required to be searched by this Authority, namely:
Remark : Although claims 14-16, 19, as far as they concern an "in vivo" method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos. because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos. because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Application No PCT/GB 94/02546

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-0415731	06-03-91	AU-B- 647747	31-03-94
		AU-A- 6199190	07-03-91
		AU-B- 6458794	08-09-94
		CN-A- 1050899	24-04-91
		JP-A- 3172189	25-07-91

WO-A-9215693	17-09-92	EP-A- 0573479	15-12-93
		JP-T- 6505156	16-06-94
